AD-A079 302 WISCONSIN UNIV-MADISON DEPT OF STATISTICS F/G 12/1
SOME NEW MATHEMATICAL METHODS FOR VARIATIONAL OBJECTIVE ANALYSI--ETC(U)
SEP 79 G WAHBA , J WENDELBERGER N00014-77-C-0675
UNCLASSIFIED UWIS-DS-578 NL

1 OF 1
AD
A079302

END
DATE
FILMED
2-80
DDC

Department of Statistics
University of Wisconsin
Madison, Wisconsin 53706

TECHNICAL REPORT NO. 578

Sep[illegible] 1979

SOME NEW MATHEMATICAL METHODS FOR VARIATIONAL OBJECTIVE ANALYSIS USING SPLINES AND CROSS-VALIDATION

by

Grace Wahba[1]

~~and~~

James Wendelberger[2]

Department of Statistics
University of Wisconsin

Typist: Debbie Dickson

[1]Research supported by the Office of Naval Research under Contract No. N00014-77-C-0675.

[2]Research supported by the National Science Foundation under Grant No. ATM75-23223

80 1 1 009

SOME NEW MATHEMATICAL METHODS FOR VARIATIONAL OBJECTIVE ANALYSIS USING SPLINES AND CROSS-VALIDATION

Grace Wahba
James Wendelberger

Department of Statistics
University of Wisconsin
Madison, Wisconsin 53706

ABSTRACT

Let $\Phi(x,y,p,t)$ be a meteorological field of interest, say height, temperature, a component of the wind field, etc. We suppose that data $\{\tilde{\Phi}_i\}_{i=1}^{N}$ concerning the field of the form $\tilde{\Phi}_i = L_i\Phi + \varepsilon_i$ are given, where each L_i is an arbitrary continuous linear functional and ε_i is a measurement error.

The data Φ_i may be the result of theory, direct measurements, remote soundings, or a combination of these. We develop a new mathematical formalism exploiting the method of Generalized Cross Validation, and some recently developed optimization results, for analyzing this data. The analyzed field, $\Phi_{N,m,\lambda}$, is the solution to the minimization problem: Find Φ in a suitable space of functions to minimize

$$\frac{1}{N}\sum_{i=1}^{N}\frac{(L_i\Phi-\tilde{\Phi}_i)^2}{\sigma_i^2} + \lambda J_m(\Phi) \qquad (*)$$

where

$$J_m(\Phi) = \sum_{\alpha_1+\alpha_2+\alpha_3+\alpha_4=m} \frac{m!}{\alpha_1!\alpha_2!\alpha_3!\alpha_4!}\iiiint\left(\frac{\partial^m\Phi}{\partial x^{\alpha_1}\partial y^{\alpha_2}\partial p^{\alpha_3}\partial t^{\alpha_4}}\right)^2 dxdydpdt\ .$$

The σ_i^2 are assumed mean square errors. Functions of d = 1,2, or 3 of the four variables x,y,p,t are also considered. Under rather general conditions, we give an explicit representation for the minimizer of (*). The parameter λ controls the tradeoff between the infidelity of the analyzed field to the data, and the roughness of the analyzed field as measured by $J_m(\cdot)$. Alternatively λ may be thought of as controlling the half-power point of the implied data filter. m controls the number of continuous derivatives that $\phi_{N,m,\lambda}$ will possess, alternatively, m may be thought of as controlling the steepness or "roll-off" of the data filter. High m corresponds to a steep roll-off. The parameters λ and m are chosen by the method of Generalized Cross Validation (GCV). This method estimates that λ and m for which the implied data filter has maximum predictive capability. This predictive capability is assessed by the GCV method by (implicitly) leaving out one data point at a time and determining how well the missing point can be predicted from the remaining data. The results extend those of Sasaki and others in several directions. In particular, no preliminary interpolation or smoothing of the data is required and it is not necessary to solve a boundary value problem or even assume boundary conditions to obtain a solution. Prior covariances are not assumed. The parameters λ and m play the role of signal to noise ratio and "order" of the covariance, these being the two most important parameters in the prior, and are estimated from the immediate data rather than historical data or guess-work. The numerical algorithm is surprisingly simple for any N with N^2 somewhat less that the high speed storage capacity of the computer.

The approach can be used to analyze temperature fields from radiosonde measured temperatures and satellite radiance measurements

simultaneously, to incorporate the geostrophic wind approximation and other information. In a test of the method (for d = 2) simulated 500mb height data was obtained at discrete points corresponding to the U.S. radiosonde network, by using an analytic representation of a 500mb wave and superimposing realistic random errors. The analytic representation was recovered on a fine grid with what appear to be impressive results.

1. Introduction

Sasaki (1960) introduced the idea of numerical variational analysis for objective analysis of meteorological fields. In the most general form of variational analysis considered here we seek a function $\phi(x,y,p,t)$ of four variables representing a meteorological field of interest, say height, temperature, or a component of the wind field, as a function of ground projection coordinates (x,y), the vertical coordinate p, and time, t. This function should be suitably close to the height, temperature or wind field as measured at a finite set of positions, pressures and times, it should reflect known behavior of such fields, and it should be "smooth" in some sense.

For an example of known behavior, fix p at 500mb, then ϕ is the 500mb geopotential height. Letting $\phi = \phi(x,y,p_0,t)$, then the sum of the tendency and horizontal advection

$$\frac{\partial\phi}{\partial t} + c_x\frac{\partial\phi}{\partial x} + c_y\frac{\partial\phi}{\partial y}$$

should be small, where c_x and c_y are the x and y components of the wind velocity. Sasaki and others have incorporated weak (i.e. approximate) and strong (i.e. exact) constraints involving the tendency, the advection, the geostrophic wind, balance, horizontal momentum, adiabatic energy, hydrostatic and continuity equations (Sasaki 1971, Lews 1972, Lewis and Grayson 1972, Achtemeier 1975).

Using the sum of the tendency and advection as a weak constraint, Sasaki (1971) suggests finding ϕ to minimize

$$J(\phi) = \iiint_R \{[\tilde{a}(\phi - \tilde{\phi})^2] + a[(\frac{\partial\phi}{\partial t} + c_x\frac{\partial\phi}{\partial x} + c_y\frac{\partial\phi}{\partial y})^2]$$

$$+ [a_t(\frac{\partial\phi}{\partial t})^2 + a_s(\frac{\partial\phi}{\partial x})^2 + a_s(\frac{\partial\phi}{\partial y})^2]\}\, dt\,dx\,dy\ , \qquad (1.1)$$

where $\tilde{a}$, a, a_t, a_s are smoothing constants to be determined, $\tilde{\phi}$ is the observed height field data, c_x and c_y are the (observed) components of wind velocity, and R is the spatial and temporal region of interest. The first term represents the desire that ϕ be close to the data, the second that the sum of the tendency and horizontal advection is small and the third that the function be "smooth" in x, y and t.

Since $\tilde{\phi}$, c_x and c_y are only measured at a (relatively sparse) set of irregularly spaced points, Sasaki assumed that the data have been interpolated to a grid sufficiently fine for numerical analytic purposes. After some simplifying assumptions, the Euler equation for the minimizer of (1.1) was obtained by Sasaki (1971) and the minimizer is found to satisfy an elliptic partial differential equation with some boundary conditions. Various authors using this and other constraints (See for example Lewis and Grayson 1972) have chosen values for the smoothing constants, and solved the resulting Euler equations numerically to obtain an objectively analyzed field.

In this paper we develop a general mathematical formalism basically embodying Sasaki's approach with four modifications:

(1) It is not necessary to first interpolate the data to a grid to obtain $\tilde{\phi}$, raw data is used directly.

(2) The problem of providing or enforcing boundary data is eliminated.

(3) The main unknown smoothing parameters are estimated from the data to be analyzed, rather than from historical data or by guesswork.

(4) The method provides a technique whereby raw indirect data, such as satellite radiance data, can be combined with direct data such as balloon temperature data in a single analysis procedure.

The method to be described avoids the problem of solving partial differential equations numerically. However it has its own challenging numerical problems, which we have been able to solve simply using existing packages for medium sized (but not large) data sets.

To introduce our general method, we begin with the simplest non-trivial example. Fix time as well as pressure and suppose that $\phi = \phi(x,y)$ is the 500mb height at (x,y) at time $t = 0$. Ignore the tendency and advection (second term) in (1.1) and suppose observations $\tilde{\phi}(x_i,y_i) = \tilde{\phi}_i$, $i = 1,2,\ldots,N$ of the 500mb height at the N stations with coordinates (x_i,y_i), $i = 1,2,\ldots,N$ are given. We want to obtain a function ϕ which is smooth and such that $\phi(x_i,y_i) \approx \tilde{\phi}_i$, $i = 1,2,\ldots,N$. Consider the minimization of

$$\frac{1}{N}\sum_{i=1}^{N} (\phi(x_i,y_i) - \tilde{\phi}_i)^2 + \lambda J_1(\phi) \qquad (1.2)$$

where

$$J_1(\phi) = \iint \left(\left(\frac{\partial\phi}{\partial x}\right)^2 + \left(\frac{\partial\phi}{\partial y}\right)^2 \right) dxdy \qquad (1.3)$$

and λ is given.

If one attempts to minimize (1.2) by e.g. writing the Euler equation one finds that the solution involves a Green's function for the Laplacian operator Δ, $\Delta\phi = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2}$, and, unfortunately, this Green's function is not bounded. Sasaki (1971) observes a similar phenomena (paragraph including (32)) but ignores it. For this and other reasons to be discussed, we seek to find the minimizer (in a suitable space of functions) of

$$\frac{1}{N}\sum_{i=1}^{N} (\phi(x_i,y_i) - \tilde{\phi}_i)^2 + \lambda J_m(\phi), \quad m = 2,3,\ldots \qquad (1.4)$$

where

$$J_2(\phi) = \iint \left(\left(\frac{\partial^2\phi}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2\phi}{\partial x\partial y}\right) + \left(\frac{\partial^2\phi}{\partial y^2}\right)^2 \right) dxdy \qquad (1.5)$$

or more generally

$$J_m(\phi) = \iint \sum_{\nu=0}^{m} \binom{m}{\nu} \left(\frac{\partial^m\phi}{\partial x^\nu \partial y^{m-\nu}}\right)^2 dxdy, \quad m = 2,3,\ldots \qquad (1.6)$$

If $J_m(\phi)$ is small, then ϕ will be "smooth".

We have deliberately omitted any mention of the domain of integration. If the domain of integration in (1.5) and (1.6) is taken as a bounded region R then it can be shown that the minimizer of (1.4) satisfies

$$\Delta^m \phi = 0, \quad (x,y) \neq (x_i,y_i), \quad i = 1,2,\ldots,N$$

where Δ is the Laplacian,

$$\Delta\phi = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2},$$

and it satisfies the natural (Neumann) boundary conditions. This result in a similar problem appears in Dyn and Wahba (1979). We avoid the necessity of solving a boundary value problem by letting the domain of integration be $-\infty \le x,y \le \infty$. Then the solution will be defined for $-\infty < x,y < \infty$. However we will only compute it on R and of course it will only have meaning if there are data points not too far from the boundary. We are also assuming here that the world is flat in R. This assumption can

be removed, see Wahba (1979c).

The solution, call it $\phi_{N,m,\lambda}$ to the problem: Find ϕ in a suitable space X to minimize

$$\frac{1}{N}\sum_{i=1}^{N}(\phi(x_i,y_i)-\tilde{\phi}_i)^2+\lambda\iint_{-\infty}^{\infty}\sum_{\nu=0}^{m}\binom{m}{\nu}\left(\frac{\partial^m\phi}{\partial x^\nu \partial y^{m-\nu}}\right)^2 dxdy \qquad (1.7)$$

was obtained by Duchon (1976a) and further studied by Meinguet (1978, 1979) and Wahba (1979a, 1979b). It is known as a "thin plate spline" and is a generalization to two dimensions of the one dimensional smoothing polynomial spline (Reinsch 1967).

We will give an explicit computable formula for $\phi_{N,m,\lambda}$ later. Problems in assigning boundary values are eliminated, and no preliminary analysis of the raw data is used.

$\phi_{N,m,\lambda}$ may be considered as the result of applying a low-pass filter to the data. In frequency space it can be shown that λ controls the half-power point of the filter and m the steepness of the roll-off. See Craven and Wahba (1979), Wahba (1978). We choose λ and m from the data by the GCV (*generalized cross-validation*) method (Craven and Wahba 1979, Golub, Heath and Wahba 1979) which proceeds as follows. The criteria for a good choice of λ and m is taken to be the ability to predict the value of the field where data is withheld.

To estimate this ability from the data let $\phi_{N,m,\lambda}^{(k)}$ be the function which is the minimizer of (1.7) with the kth data point omitted. If λ and m are good choices, then on the average $\phi_{N,m,\lambda}^{(k)}(x_k,y_k)-\tilde{\phi}_k$ should be small and we measure this by the "ordinary cross-validation function"

$$V_m^*(\lambda)\equiv\frac{1}{N}\sum_{k=1}^{N}(\phi_{N,m,\lambda}^{(k)}(x_k,y_k)-\tilde{\phi}_k)^2 . \qquad (1.8)$$

This expression is rather nasty to compute, furthermore effect of unequal spacing of data points are not suitably accounted for. For these and other technical reasons recounted in Craven and Wahba (1979) and Golub, Heath and Wahba (1977), one should measure the ability of $\phi_{N,m,\lambda}$ to predict missing data by the "generalized cross-validation function" (GCVF)

$$V_m(\lambda)=\frac{1}{N}\sum_{k=1}^{N}(\phi_{N,m,\lambda}^{(k)}(x_k,y_k)-\tilde{\phi}_k)^2\, w_k(m,\lambda) \qquad (1.9)$$

where the $w_k(m,\lambda)$ are certain weights which have been given in Craven and Wahba (1979) and Golub, Heath and Wahba (1979). $V_m(\lambda)$ turns out to have a collapsed representation which is relatively easy to compute. For each m = 2,3,4..., up to some preset maximum, $V_m(\lambda)$ is computed as a function of λ and the [illegible] $\hat{\lambda}(m)$ of λ minimizing $V_m(\lambda)$ is determined. Then m is selected by comparing $V_m(\hat{\lambda}(m))$ over m. A computer implementation of this example has been made and applied to data simulated from a mathematical model for a 500mb height field. The results are presented in Section 4.

We next generalize this approach to allow the imposition of weak constraints. Continuing with p = 500mb, t = t_0, we consider as an example the geostrophic wind approximation:

$$u_g \approx \frac{-1}{f}\frac{\partial\phi}{\partial y}$$

$$v_g \approx \frac{1}{f}\frac{\partial\phi}{\partial x}$$

where ϕ is the 500mb height, u_g and v_g are eastward and northward components of the geostrophic wind, and f is the Coriolis parameter. If the eastward and northward components of the wind are measured at each station, one can seek ϕ to minimize

$$\frac{1}{N}\sum_{i=1}^{n}\frac{1}{\sigma_1^2}\left(\phi(x_i,y_i)-\tilde{\phi}_i\right)^2+\frac{1}{N}\sum_{i=1}^{n}\frac{1}{\sigma_2^2}\left(\frac{\partial\phi}{\partial y}\bigg|_{x_i,y_i}+f\,\tilde{u}_i\right)^2$$

$$+\frac{1}{N}\sum_{i=1}^{n}\frac{1}{\sigma_3^2}\left(\frac{\partial\phi}{\partial x}\bigg|_{x_i,y_i}-f\,\tilde{v}_i\right)^2+\lambda J_m(\phi) \qquad (1.10)$$

where $N = 3n$, $\tilde{\phi}_i$ is the measured 500mb height and $\tilde{u}_i, \tilde{v}_i$ are the observed wind components at station i. σ_1^2 is a weight which is ideally, the mean square error in the measured height field. σ_2^2 is the sum of the mean square error in the measured eastward component of the wind and the mean square error in the geostrophic approximation to the true eastward wind. σ_3^2 has the corresponding meaning for the northward component of the wind.

For $m \geq 3$ an explicit formula for the minimizer $\phi_{N,m,\lambda}$ of (1.10) will be given.

Since we are going to choose λ from the data, it is only necessary that σ_1^2/σ_2^2 and σ_1^2/σ_3^2 are known reasonably well. Assuming all mean square errors are known, it has been suggested by Reinsch (1967) and others to choose λ so that the first three terms in (1.10) with ϕ replaced by $\phi_{N,m,\lambda}$ sum to 1. It has been shown, however (see Wahba 1975, Craven and Wahba 1979) that this will lead systematically to undersmoothing.

The idea of the generalized cross-validation function extends to the choice of λ and m in the minimizer of (1.10) and we can obtain the GCVF $V_m(\lambda)$ which can be minimized to estimate good values of m and λ.

In this example where σ_1^2, σ_2^2 and σ_3^2 may be different the minimizer of the GCVF estimates λ and m which best predict missing data points, inversely weighted by the appropriate σ_i^2.

We next turn to the analysis of a temperature field using both direct (balloon) and remote (satellite radiance) data. We assume that

all data are measured at $t = 0$ and that $\phi(x,y,p)$ represents the temperature. The data consists of direct measurement of the temperature from station i at pressure p_i, and indirect satellite measurements of radiances $I_\ell(\nu)$ at frequency ν and subsatellite point (x_ℓ, y_ℓ). In the simplest case (cloudless, looking down), after some linearization and approximations[2], a known function $r_\ell(\nu)$ of the measured radiance $I_\ell(\nu)$ can be related to the temperature ϕ by

$$r_\ell(\nu) = \int_0^{P_o} K(\nu,p)\,\phi\,(x_\ell,y_\ell,p)\,dp \qquad (1.11)$$

where $K(\nu,p)$ is known for each frequency $\nu = \nu_1,\ldots,\nu_n$. (See Fritz et al (1972)).

Thus we seek ϕ to minimize

$$\frac{1}{N}\sum_{i,k}\frac{1}{\sigma_k^2}\left(\phi(x_i,y_i,p_k)-\tilde{\phi}_{ik}\right)^2$$

$$+\frac{1}{N}\sum_{\ell,\nu}\frac{1}{\sigma_\nu^2}\left(\int_0^{P_o}K(\nu,p)\,\phi\,(x_\ell,y_\ell,p)\,dp-r_\ell(\nu)\right)^2$$

$$+\lambda J_m(\phi) \qquad (1.12)$$

where N is the total number of observations and

2/ In the process of linearizing to obtain (1.11) a "first guess" field is used. One could obtain this first guess field by analyzing the balloon data alone by leaving out the radiance data in what follows. (Second term in (1.12)).

$$J_m(\phi) = \sum_{\alpha_1+\alpha_2+\alpha_3=m} \left(\frac{m!}{\alpha_1!\alpha_2!\alpha_3!}\right) \iiint_{-\infty}^{\infty} \left(\frac{\partial^m \phi}{\partial x^{\alpha_1} \partial y^{\alpha_2} \partial p^{\alpha_3}}\right)^2 dx\,dy\,dp \tag{1.13}$$

We will give an explicit formula for the minimizer $\phi_{N,m,\lambda}$, of (1.12) and the GCVF $V_m(\lambda)$ for this problem for $m \geq 2$. In theory, there is no difficulty in adding weak temperature constraints, or in carrying out the analysis in 3 space variables and one time variable with direct data, indirect data and weak constraints. (A finite number of strong constraints can be added too, we briefly indicate how.) In practice, the method has computational limits. The computation of $\phi_{N,m,\lambda}$ requires the solution of a linear system of dimension close to the number N of "data" and "weak constraint" terms. The computation of the GCVF required the solution of an eigenvalue problem of size N. We are obtaining very good results with N up to as large as 140 with present methods on the Univac 1108 at the University of Wisconsin, Madison, but improved algorithms will have to be developed to go beyond this point on this size machine. There is reason to believe that this can be done. Some algorithms handling four times as many points in certain special cases have been developed by Paihua (1978).

In Section 2 we solve a general minimization problem of which all the previously mentioned problems are special cases. In Section 3 we give the GCVF and discuss computational methods. In Section 4, results of a Monte Carlo test of the method is given, using realistic simulated 500mb height data where the "true" field is known.

Analysis of the height field via minimization of (1.4) is an anisotropic method. Thiebaux (1977) has provided some evidence that an improved analysis may be obtained using methods which have different

north-south and east-west scales. This feature may be incorporated here by making a change of scale $x \to kx$ and $y \to ky^{-1}$. A good scale parameter k may be estimated by GCV simultaneously with λ and m. Some very preliminary numerical results with actual reported 500mb height data from the U.S. rawinsonde network suggests that the $k = 1$ (i.e. isotropic) analysis can be improved upon by estimating k. See Wendelberger (1980). We do not discuss anisotropic methods further here.

Kreiss (1979a, b) notes that for successful numerical solution of certain differential equations related to numerical weather forecasting, it is desireable to have initial conditions that have certain continuity properties. We conjecture that the methods suggested here can be used to provide these initial conditions.

2. Solution of a General Minimization Problem

In this Section we give a solution to a general minimization problem of which the minimization problems of (1.7), (1.10) and (1.12) are special cases.

Our results hold in any number of dimensions, most meteorological problems of interest will involve d = 2,3 or 4. The d = 1 case is the familiar polynomial smoothing spline, See Reinsch (1967). We will say a function u of d variables $x_1, x_2, \ldots, x_d$ is "smooth" if $J_m(u)$ defined by

$$J_m(u) = \sum_{\alpha_1+\ldots+\alpha_d=m} \frac{m!}{\alpha_1!\alpha_2!\ldots\alpha_d!} \int\cdots\int_{R^d} \left(\frac{\partial^m u}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}} \right)^2 dx_1 \ldots dx_d \ , \tag{2.1}$$

is small. We will seek a "smooth" solution to our general minimization problem in a suitable (Hilbert) space $\mathcal{X}$ of functions (which depends on d and m) for which $J_m(u)$ is well defined and finite. ($\mathcal{X}$ is defined rigorously in the mathematical appendix.)

In this section we will state our mathematical results without proof. Most of the theory has appeared elsewhere, principally in the papers of Duchon (1976a), Meinguet (1978) and Wahba (1979a). In the mathematical appendix we will sketch the proof, relying heavily on previously published work. Minimization problems like those of (1.7), (1.10) and (1.12) can be solved explicitly when the data and the weak constraints are expressible in terms of continuous linear

functionals on a (Hilbert) space $\mathcal{X}$ of functions u for which $J_m(u) < \infty$. A (real) linear functional L on a Hilbert space is a functional which assigns a real number to each function u in the space with the property

$$Lu_1 + Lu_2 = L(u_1 + u_2) \ .$$

(See Akhiezer and Glazman (1961) for a basic introduction to Hilbert spaces, hopefully it is not necessary to understand Hilbert spaces to follow our results, only the proofs.) A <u>continuous</u> linear functional L in a Hilbert space is a linear functional with the following property: There exits a constant c depending only on L such that for each u in the space

$$|Lu| \le c\|u\|$$

where $\|u\|$ is the norm of u in the space. (We will shortly give examples.)

We will identify a norm suitable for our purposes after making some observations about the special role of polynomials with respect to the smoothness criteria, J_m. In d dimensional space there are $M = \binom{d+m-1}{d}$ polynomials of total degree less than or equal to m-1. Letting $\{\phi_\nu\}_{\nu=1}^{M}$ be these polynomials we have for example, for d=2, m=2, then M=3 and,

$$\phi_1(x_1,x_2) = 1, \quad \phi_2(x_1,x_2) = x_1 \quad \phi_3(x_1,x_2) = x_2 \ .$$

Observe that $J_m(\phi_\nu) = 0$, $\nu = 1,2,\ldots,M$, so that polynomials of total degree $\leq m-1$ are infinitely "smooth".

We can construct a proper norm on $\mathcal{X}$ from J_m if something is added to guarantee that the norm of ϕ_ν is not 0. A suitable norm for our purposes is given by

$$\|u\|_{\mathcal{X}} = \left(J_m(u) + \sum_{j=1}^{M} u^2(\underline{s}_j)\right)^{1/2} \tag{2.2}$$

where $\underline{s}_j = (x_1^j, x_2^j, \ldots, x_d^j)$, $j=1,2,\ldots,M$ are M points in d-space such that

$$\sum_{j=1}^{M} \phi_\nu^2(\underline{s}_j) > 0$$

for each polynomial ϕ_ν, $\nu = 1,2,\ldots,M$. The particular choice of the $\underline{s}_j$ is irrevelant here and will cancel out in the calculation of our solution to the general minimization problem below. With the norm defined by (2.2) it can be shown (Duchon, 1977) that

$$Lu = u(\underline{t}^*)$$

is a continuous linear functional on $\mathcal{X}$ for each fixed $\underline{t}^* = (x_1^*, \ldots, x_d^*)$ provided

$$2m - d > 0 \tag{2.3}$$

and that

$$Lu = \left.\frac{\partial^k u}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}\right|_{(x_1,\ldots,x_d) = \underline{t}^*}$$

with

$$\alpha_1 + \ldots + \alpha_d = k$$

is a continuous linear functional on $\mathcal{X}$ for each fixed $\underline{t}^*$, provided that

$$2m - 2k - d > 0 \quad . \tag{2.4}$$

Linear functionals of the form

$$Lu = \int\ldots\int_\Omega K(x_1,\ldots,x_d) u(x_1,\ldots,x_d)\, dx_1 \ldots dx_d$$

are also continuous on $\mathcal{X}$ if for example Ω is a bounded set and $\int\ldots\int_\Omega |K(x_1,\ldots,x_d)| dx_1 \ldots dx_d < \infty$. We remark that $Lu = u(\underline{t}^*)$ is not a continuous linear functional on $\mathcal{X}$ if $m = 1$, $d = 2$, and this leads to the difficulties mentioned previously in regard to the minimization of (1.2).

The reason one can explicitly solve optimization problems of the type we are considering when the data functionals are continuous linear functionals is that for each such continuous linear functional L we are guaranteed the existence of a function $\eta \in \mathcal{X}$, to be called the representer of L, with the property that

$$Lu = \langle \eta, u \rangle_{\mathcal{X}} \text{ for every } u \in \mathcal{X}.$$

where $\langle \cdot \rangle_{\mathcal{X}}$ is the inner product in $\mathcal{X}$. (See Akhiezer and Glazman (1961)). Then it is known (See, for example Kimeldorf and Wahba (1971)) that the solution to the problem:

Find $u \in \mathcal{X}$ to minimize:

$$\frac{1}{N} \sum_{k=1}^{N} \left(\frac{L_k u - z_k}{\sigma_k} \right)^2 + \lambda J_m(u) \qquad (2.5)$$

where $\{z_k\}$ are given numbers, is of necessity a linear combination of the representers $\{\eta_k\}_{k=1}^{N}$ of the $\{L_k\}_{k=1}^{N}$ and the set of functions for which J_m is 0. This latter set is the M polynomials $\{\phi_\nu\}_{\nu=1}^{M}$ of total degree $\leq m-1$. It is necessary to find the representers $\eta_k = \eta_k(x_1, \ldots, x_d)$ of the L_k to give the solution explicitly. Suppose the linear functional $L_{\underset{\sim}{t}}$ defined by $L_{\underset{\sim}{t}} u = u(\underset{\sim}{t})$ is a continuous linear functional in $\mathcal{X}$ for each fixed $\underset{\sim}{t}$. Then $\mathcal{X}$ is called a reproducing kernel space (See Aronszajn (1950)) and one can define a reproducing kernel $K(\underset{\sim}{s}, \underset{\sim}{t})$, which is a symmetric function of the two vector variables $\underset{\sim}{s}, \underset{\sim}{t}$ by

$$K(\underset{\sim}{s}, \underset{\sim}{t}) = \langle \eta_{\underset{\sim}{s}}, \eta_{\underset{\sim}{t}} \rangle_{\mathcal{X}}$$

where $\eta_{\underset{\sim}{t}}$ is the representer for $L_{\underset{\sim}{t}}$. It is well known (see, e.g. Kimeldorf and Wahba (1971) and references there) that, if L is any continuous linear functional then its representer η is given by the formula

$$\eta(\underset{\sim}{t}) = L_{(\underset{\sim}{s})} K(\underset{\sim}{t}, \underset{\sim}{s})$$

where $L_{(\underset{\sim}{s})} K(\underset{\sim}{t}, \underset{\sim}{s})$ means L applied to K considered as a function of $\underset{\sim}{s}$. The reproducing kernel for the space $\mathcal{X}$ with norm given by (2.2) is given in the mathematical appendix.

Given this reproducing kernel one can deduce the following Theorem:

Let $L_1, L_2, \ldots, L_N$ be N linearly independent continuous linear functionals on $\mathcal{X}$ and suppose

$$L_k \sum_{\nu=1}^{M} a_\nu \phi_\nu = 0, \; k = 1, 2, \ldots, N \qquad (2.6)$$

implies that all the a_ν are 0.

Then the solution to the problem: Find $u_{N,m,\lambda} \in \mathcal{X}$ to minimize

$$\frac{1}{N} \sum_{j=1}^{N} \left(\frac{L_j u - z_j}{\sigma_j} \right)^2 + \lambda J_m(u) \qquad (2.7)$$

is unique and has the representation

$$u_{N,m,\lambda}(\underset{\sim}{t}) = \sum_{j=1}^{N} c_j \xi_j(\underset{\sim}{t}) + \sum_{\nu=1}^{M} d_\nu \phi_\nu(\underset{\sim}{t}) \qquad (2.8)$$

where

$$\underset{\sim}{t} = (x_1, x_2, \ldots, x_d)$$

$$\xi_j(\underset{\sim}{t}) = L_{j(\underset{\sim}{s})} E_m(\underset{\sim}{t}, \underset{\sim}{s}) \qquad j = 1, 2, \ldots, N \qquad (2.9)$$

$$\phi_\nu(\underset{\sim}{t}) = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_d^{\alpha_d} \qquad \nu = 1, 2, \ldots, M \qquad (2.10)$$

where each ν stands for one of the M sets $(\alpha_1,\ldots,\alpha_d)$ with $\sum_{i=1}^{d} \alpha_i \le m-1$. The function $E_m(\underset{\sim}{s},\underset{\sim}{t})$ of two vector variables $(\underset{\sim}{s},\underset{\sim}{t})$ is defined as follows. If $\underset{\sim}{s} = (y_1,\ldots,y_d)$, $\underset{\sim}{t} = (x_1,\ldots,x_d)$, then

$$|\underset{\sim}{s}-\underset{\sim}{t}| = \left(\sum_{i=1}^{d} (x_i-y_i)^2\right)^{1/2} \text{, and} \tag{2.11}$$

$$E_m(\underset{\sim}{s},\underset{\sim}{t}) = E(|\underset{\sim}{s}-\underset{\sim}{t}|) \tag{2.12}$$

where

$$\begin{aligned} E(\tau) &= \theta_m \tau^{2m-d} \ln \tau && d \text{ even} \\ &= \theta_m \tau^{2m-d} && d \text{ odd} \end{aligned}$$

where

$$\begin{aligned} \theta_m &= \frac{(-1)^{d/2+1+m}}{2^{2m-1}\pi^{d/2}(m-1)!(m-d/2)!} && d \text{ even} \\ &= \frac{\Gamma(d/2-m)}{2^{2m}\pi^{d/2}(m-1)!} && d \text{ odd} . \end{aligned}$$

The notation $L_{j(\underset{\sim}{s})}E_m(\underset{\sim}{s},\underset{\sim}{t})$ means that the linear functional L_j is applied to E_m considered as a function of $\underset{\sim}{s}$. The coefficient vectors $\underset{\sim}{c} = (c_1,\ldots,c_N)$ and $\underset{\sim}{d} = (d_1,\ldots,d_M)$ are determined by

$$(K+N\lambda D_\sigma^2)\underset{\sim}{c} + T\underset{\sim}{d} = \underset{\sim}{z} \tag{2.13}$$

$$T'\underset{\sim}{c} = 0 \tag{2.14}$$

where K is the N x N symmetric matrix with jk^{th} entry

$$L_{j(\underset{\sim}{s})}L_{k(\underset{\sim}{t})} \quad E_m(\underset{\sim}{s},\underset{\sim}{t}) \quad . \tag{2.15}$$

T is the N x M matrix with $j\nu^{th}$ entry

$$L_j\phi_\nu \tag{2.16}$$

and D_σ is the N x N diagonal matrix with jj^{th} entry σ_j.

Example: The simplest example is when the bounded linear functionals are all evaluation functionals: $L_i u = u(\underset{\sim}{t}_i)$, $i = 1,2,\ldots,N$. For condition (2.6) to be satisfied it is necessary that the N points $\underset{\sim}{t}_1,\ldots,\underset{\sim}{t}_N$ do not lie in a hyperplane of dimension d-1 or less. For example, if $d = 2$, then we need $N \ge \binom{m+1}{2}$ and the N points must not fall on a straight line. Then

$$L_{j(\underset{\sim}{s})} \; E_m(\underset{\sim}{s},\underset{\sim}{t}) = E_m(\underset{\sim}{t}_j,\underset{\sim}{t}) = \xi_j(\underset{\sim}{t}) \tag{2.17}$$

$$L_{j(\underset{\sim}{s})}L_{k(\underset{\sim}{t})}E_m(\underset{\sim}{s},\underset{\sim}{t}) = E_m(\underset{\sim}{t}_j,\underset{\sim}{t}_k) \tag{2.18}$$

$$L_j\phi_\nu = \phi_\nu(\underset{\sim}{t}_j) \; . \tag{2.19}$$

If $$Lu = \left.\frac{\partial^k u}{\partial x_1^{\alpha_1}\cdots\partial x_d^{\alpha_d}}\right|_{(x_1,\ldots,x_d)=(x_1^*,\ldots,x_d^*)}, \quad k = \alpha_1+\cdots+\alpha_d ,$$

we have

$$\xi(x_1,\ldots,x_d) = \left.\frac{\partial^k}{\partial y_1^{\alpha_1}\cdots\partial y_d^{\alpha_d}} E_m(y_1,\ldots,y_d;x_1,\ldots,x_d)\right|_{(y_1,\ldots,y_d)=(x_1^*,\ldots,x_d^*)}$$

where

$$E_m(y_1,\ldots,y_d;x_1,\ldots,x_d) = E_m(\underset{\sim}{s},\underset{\sim}{t})$$

with

$$\underset{\sim}{t} = (x_1,\ldots,x_d)$$

$$\underset{\sim}{s} = (y_1,\ldots,y_d) \quad .$$

The differentiation can be carried out explicitly and the ξ's, K and T found explicitly. For example, if $d = 2$, $m = 3$, $L_j u = \left.\frac{\partial}{\partial x_1} u\right|_{(x_1^j,x_2^j)}$, $L_k u = \left.\frac{\partial}{\partial x_1} u\right|_{(x_1^k,x_2^k)}$, say, then $\xi_j(x_1,x_2) = \left.\frac{\partial}{\partial y_1} E_3(y_1,x_2^j;x_1,x_2)\right|_{y_1=x_1^j}$

$$= \left.\frac{\partial}{\partial y_1}\frac{\theta_3}{2}[(y_1-x_1)^2+(x_2^j-x_2)^2]^2 \ln[(y_1-x_1)^2+(x_2^j-x_2)^2]\right|_{y_1=x_1^j}$$

$$= \theta_3\{2[(x_1^j-x_1)^2+(x_2^j-x_2)^2](x_1^j-x_1)\ln[(x_1^j-x_1)^2+(x_2^j-x_2)^2]$$

$$+ [(x_1^j-x_1)^2+(x_2^j-x_2)^2](x_1^j-x_1)\} .$$

$$L_{j(\underset{\sim}{s})} L_{k(\underset{\sim}{t})} E_m(\underset{\sim}{s},\underset{\sim}{t}) = \frac{\partial^2}{\partial y_1 \partial x_1} \quad \frac{\theta_3}{2}[(y_1-x_1)^2+(x_2^j-x_2^k)^2] \times$$

$$\left.\ln[(y_1-x_1)^2+(x_2^j-x_2^k)^2]\right|_{\substack{y_1=x_1^j \\ x_1=x_1^k}}$$

$$= \left.\frac{\partial}{\partial x_1}\xi_j(x_1,x_2^k)\right|_{x_1=x_1^k} .$$

etc.

If

$$Lu = \int_a^b K(x_3)u(x_1^*,x_2^*,x_3)\,dx_3$$

then

$$\xi(x_1,x_2,x_3) = \int_a^b K(y_3)E_m(x_1^*,x_2^*,y_3;x_1,x_2,x_3)dy_3 \quad ,$$

etc. In general ξ of this form may not be known explicitly, then a quadrature approximation may be necessary. An appropriate quadrature approximation for a similar problem can be found in Dyn and Wahba (1979). We discuss the calculation of the coeffecient vectors $\underset{\sim}{c}$ and $\underset{\sim}{d}$ of (2.13) and (2.14) in conjunction with the calculation of the GCVF in the next section.

We remark that in the more familiar Hilbert spaces of functions for which

$$\|u\| = \left(\int \cdots \int u^2(x_1,\ldots,x_d)\,dx_1,\ldots,dx_d\right)^{1/2} ,$$

the evaluation functionals $L_k u = u(\underset{\sim}{t}_k)$ are not continuous.

3. Cross Validation and the GCVF

Before discussing cross validation we remark that the requirement to compute the GCVF will influence the choice of method of solving (2.13) and (2.14) for $\underset{\sim}{c}$ and $\underset{\sim}{d}$.

We first define the "ordinary" cross validation function $V_m^o(\lambda)$. Let $u_{N,m,\lambda}^{(k)}$ be the minimizer of

$$\frac{1}{N} \sum_{\substack{j=1 \\ j \neq k}}^{N} \frac{(L_j u - z_j)^2}{\sigma_j^2} + \lambda J_m(u) .$$

Then,

$$L_k\, u_{N,m,\lambda}^{(k)} - z_k \tag{3.1}$$

is the difference between the j^{th} data point and an estimate of the j^{th} data point from the remaining data when m and λ are used. If m and λ are a good choice the quantities in (3.1) should be small on the average and we measure this by

$$V_m^o(\lambda) = \frac{1}{N} \sum_{k=1}^{N} \frac{(L_k u_{N,m,\lambda}^{(k)} - z_k)^2}{\sigma_k^2} . \tag{3.2}$$

A simplified expression for $V_m^o(\lambda)$ can be obtained in terms of a certain $N \times N$ matrix $A_m(\lambda)$ defined by

$$\begin{pmatrix} L_1 u_{N,m,\lambda} \\ L_2 u_{N,m,\lambda} \\ \vdots \\ L_N u_{N,m,\lambda} \end{pmatrix} = A_m(\lambda)\underset{\sim}{z} \qquad (3.3)$$

where $\underset{\sim}{z} = (z_1,\ldots,z_N)'$. A formula for $A_m(\lambda)$ will be given later. $A_m(\lambda)$ is to be thought of as the matrix which maps the data vector $\underset{\sim}{z}$ into the "smoothed data vector" $(L_1 u_{N,m,\lambda}, L_2 u_{N,m,\lambda},\ldots,L_N u_{N,m,\lambda})'$. This simplified expression for $V_m^*(\lambda)$ can be obtained through the use of the following.

Theorem:

$$L_k u_{N,m,\lambda}^{(k)} - z_k = (L_k u_{N,m,\lambda} - z_k)/(1-a_{kk}) \quad , \qquad (3.4)$$

where $a_{kk} = a_{kk}(m,\lambda)$ is the kk^{th} entry of $A_m(\lambda)$.

<u>Proof</u>. This theorem is proved for the special case $d = 1$, $\sigma_i^2 = 1$ and $L_k u = u(t_k)$ in Craven and Wahba (1979). However the argument in Craven and Wahba is a variational one independent of the context and the proof of present theorem follows immediately. We have as a result that $V_m^*(\lambda)$ of (3.2) is also given by

$$V_m^*(\lambda) \equiv \frac{1}{N}\sum_{k=1}^{N}\frac{1}{\sigma_k^2}\frac{\left(L_k u_{N,m,\lambda} - z_k\right)^2}{(1-a_{kk})^2} \quad . \qquad (3.5)$$

We are assuming that the criteria for choosing λ and m is

based on the following model:

$$z_j = L_j u + \epsilon_j \ , \quad j = 1,2,\ldots,N$$

where u is the "true" field and ϵ_j is an error which is assumed to have mean 0 and mean square σ_j^2. We define an error function when m and λ are used as

$$R_m(\lambda) = E\,\frac{1}{N}\sum_{j=1}^{N}\frac{\left(L_j u - L_j u_{N,m,\lambda}\right)^2}{\sigma_j^2}, \qquad (3.6)$$

where the E means "expected value". $R_m(\lambda)$ is not computable of course, since u is not known. It would be nice to establish that λ and m which minimize $V_m^*(\lambda)$ are good estimates of the λ and m which minimize $R_m(\lambda)$. It has been shown under some moderately general assumptions on $A_m(\lambda)$ which will usually be satisfied here (see Golub, Heath and Wahba 1977), that it is better to minimize the GCVF $V_m(\lambda)$ defined by

$$V_m(\lambda) = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{\sigma_k^2}\left(L_k u_{N,m,\lambda}^{(k)} - z_k\right)^2 w_k(m,\lambda) \qquad (3.7)$$

$$= \frac{1}{N}\sum_{k=1}^{N}\frac{1}{\sigma_k^2}\left(L_k u_{N,m,\lambda} - z_k\right)^2 \ w_k(m,\lambda)/(1 - a_{kk}) \qquad (3.8)$$

where

$$w_k = (1 - a_{kk})^2/(1 - \frac{1}{N}\sum_{i=1}^{N} a_{ii})^2.$$

By better, means, on the average to better estimate the minimizers of $R_m(\lambda)$. In fact, under certain general conditions for large N, $EV_m(\lambda) \simeq R_m(\lambda) + 1$, see Craven and Wahba (1979). In the special case $a_{11} = a_{22} = \dots = a_{NN}$ then $V_m(\lambda)$ and $V_m^o(\lambda)$ coincide. It is fortunate that it is better to minimize $V_m(\lambda)$ because $V_m(\lambda)$ is much easier to compute than $V_m^o(\lambda)$. In particular using the definition of $A_m(\lambda)$ in (3.3) it follows that

$$V_m(\lambda) \equiv \frac{\frac{1}{N}\|D_\sigma^{-1}(I-A_m(\lambda))z\|^2}{[\frac{1}{N}\,\mathrm{Trace}\,(I-A_m(\lambda))]^2} \tag{3.9}$$

where D_σ is the diagonal matrix with jj^{th} entry σ_j, the trace of a matrix is the sum of its diagonal entries, and $\|\cdot\|$ is the Euclidean norm.

The following facts are established in the mathematical appendix:

i) $\underset{\sim}{c} = R(R'KR + N\lambda R'D_\sigma^2R)^{-1}R'z$ (3.10)

ii) $\underset{\sim}{d} = (T'D_\sigma^{-2}T)^{-1}\,T'D_\sigma^{-2}(\underset{\sim}{z} - K\underset{\sim}{c})$ (3.11)

($\underset{\sim}{c}$ and $\underset{\sim}{d}$ have originally been given in (2.13) and (2.14)).

iii) $I-A_m(\lambda) = N\lambda D_\sigma^2R(R'KR + N\lambda R'D_\sigma^2R)\,R'$ (3.12)

where R is any $N \times N\text{-}M$ dimensional matrix of rank N-M satisfying $R'T=0_{N-M \times M}$.

iv) The $N\text{-}M \times N\text{-}M$ dimensional matrix B defined by $B = R'KR$ is always strictly positive definite (although K may not be).

We now discuss a computational procedure which we have successfully implemented for the special case $d = 2$, $L_ju = u(\underset{\sim}{t}_j)$, $\sigma_j^2 \equiv \sigma^2$, $m = 2,3,4,5$, or 6, and $N \le 120$.

R can always be chosen so that $R'D_\sigma^2R = I_{N-M}$ where I_{N-M} is the N-M dimensional identity matrix. This is done numerically as follows: Let $\tilde{T} = D_\sigma^{-1}T$ and form the matrix $C=I-\tilde{T}(\tilde{T}'\tilde{T})^{-1}\tilde{T}'$. This symmetric non negative definite matrix is a projection matrix of rank N-M satisfying $\tilde{T}'C = 0_{M \times N}$, and so it has N-M eigenvalues equal to one and M eigenvalues equal to 0. The N-M eigenvectors $\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_{N-M}$, say, corresponding to the ones have the property

$$\tilde{T}'\tilde{r}_j = 0,\ j = 1,2,\dots,N\text{-}M \ ,$$

and the property that the $N \times N\text{-}M$ dimensional matrix $\tilde{R}$ with columns $\tilde{r}_1, \dots, \tilde{r}_{N-M}$ satisfies $\tilde{R}'\tilde{R} = I$. The eigenvectors corresponding to the ones are not individually uniquely defined of course, any set will do. Let $r_j = D_\sigma^{-1}\tilde{r}_j$ and $R = D_\sigma^{-1}\tilde{R}$. Then $T'r_j = \tilde{T}'\tilde{r}_j = 0$, $j = 1,2,\dots,N\text{-}M$ and $R'D_\sigma^2R = \tilde{R}'\tilde{R} = I$. Thus R is the desired matrix. We successfully used EISPACK (Smith et al (1976)) in double precision to deliver the $\{\tilde{r}_j\}$ given C, for N up to about 120. Once R is determined, let the eigenvalue decomposition of $B = RKR'$ be

$$B = UD_BU'$$

where U is orthogonal and D_B is the diagonal matrix with diagonal entries the eigenvalues b_i, of B, $i = 1,2,\ldots,N-M$. The b_i are theoretically all positive. Then $\underset{\sim}{c}$ is readily computed from the identity

$$\underset{\sim}{c} = RU'(D_B + N\lambda I)^{-1}UR'\underset{\sim}{z} \quad .$$

and it is shown in the appendix that

$$\underset{\sim}{d} = (\tilde{T}'\tilde{T})^{-1}\tilde{T}'D_\sigma^{-1}(\underset{\sim}{z} - K\underset{\sim}{c}) \quad .$$

Since

$$D_\sigma^{-1}(I - A_m)\underset{\sim}{z} = N\lambda D_\sigma \underset{\sim}{c}$$

we have

$$V_m(\lambda) = \frac{\frac{1}{N}(N\lambda)^2 \| D_\sigma RU'(D_B + N\lambda I)^{-1}\underset{\sim}{w}\|^2}{\frac{1}{N^2}(N\lambda)^2(\text{Trace } R'D_\sigma^{\,2}R(B + N\lambda R'D_\sigma^{\,2}R)^{-1})^2}$$

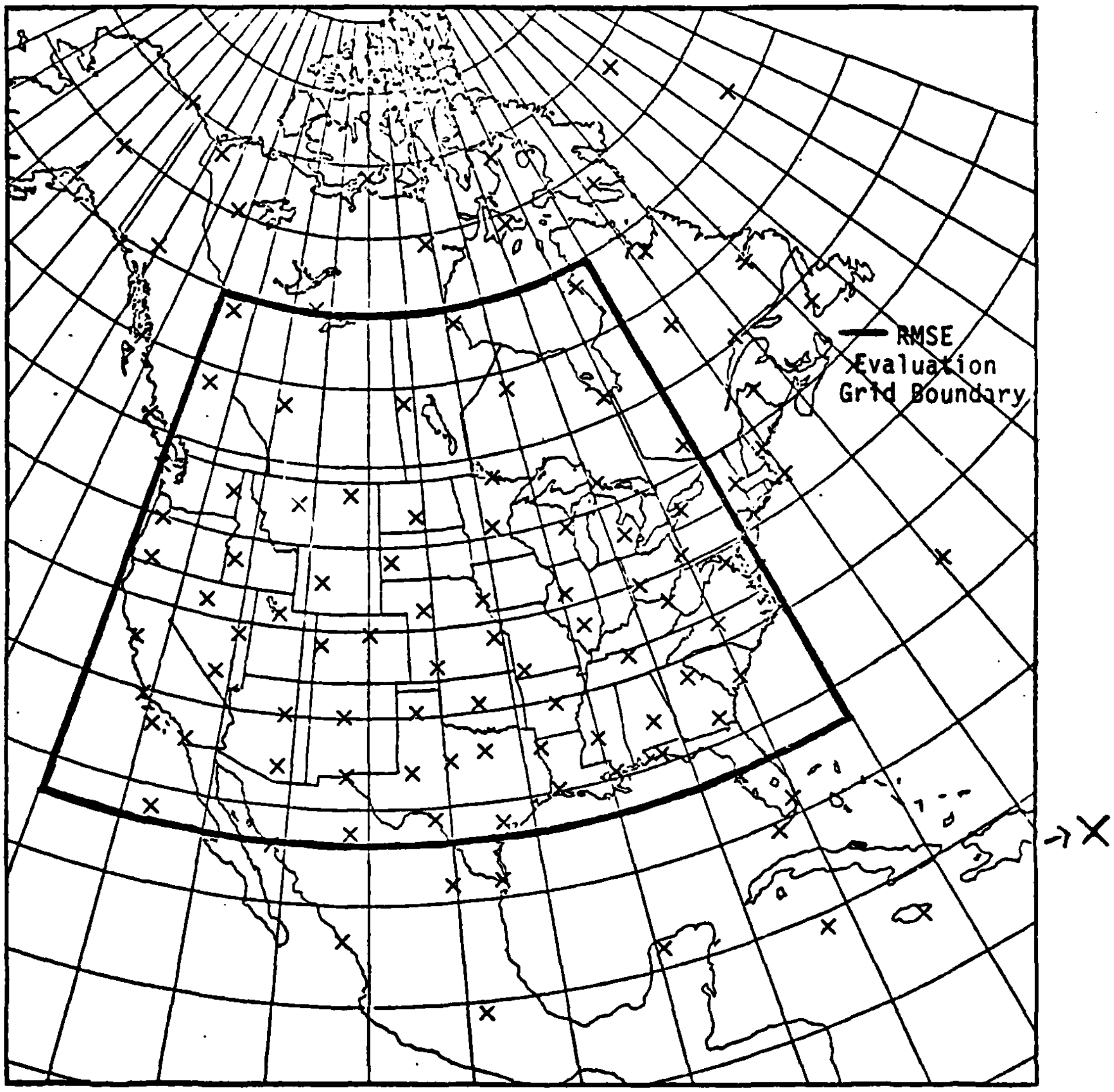
$$= \frac{\frac{1}{N}\sum_{i=1}^{N-M}\frac{w_i^2}{(b_i + N\lambda)^2}}{\left(\frac{1}{N}\sum_{i=1}^{N-M}\frac{1}{b_i + N\lambda}\right)^2}$$

where $\underset{\sim}{w} = (w_1,\ldots,w_{N-M})' = UR'\underset{\sim}{z}$.

4. Numerical Experiments

We have programmed and tested the method for analyzing data from simulated 500mb height fields using simulated data at N = 114 North American radiosonde station locations. The simulated data were obtained from a mathematical model of 500mb height fields used by Dr. Thomas Koehler of the Department of Meteorology at the University of Wisconsin that was based on an earlier model developed by Sanders (1971). The location of the 114 stations is given in Figure 1. The equations generating the field are given in Appendix 2. Discussion of the rationale behind the model appears in Koehler's thesis (1979). Contour maps of the model fields appear below together with contour maps of the analyzed fields determined from the simulated data. Data were simulated by computing the "true" 500mb height at station i by calling Koehler's program and adding a simulated measurement error. The simulated measurement error was obtained by calling the pseudo random number generator RAENBR in the University of Wisconsin Academic Computing Center library. This program obtains a pseudo random normally distributed number with mean 0 and standard deviation 1 and mutliplies this number by a constant which is given here as the standard deviation of the measurement error. This procedure resulted in a set of 114 simulated measured 500mb heights which were then used to obtain an analyzed field. This is the simulated data vector $\underset{\sim}{z}$. To recapitulate the formulas for obtaining the analyzed field, we go back to the main theorem. Here N = 114, d = 2, and we have

FIGURE 1



Location of Model Radiosonde Stations and Boundary of Grid used for Evaluation of the Analysis

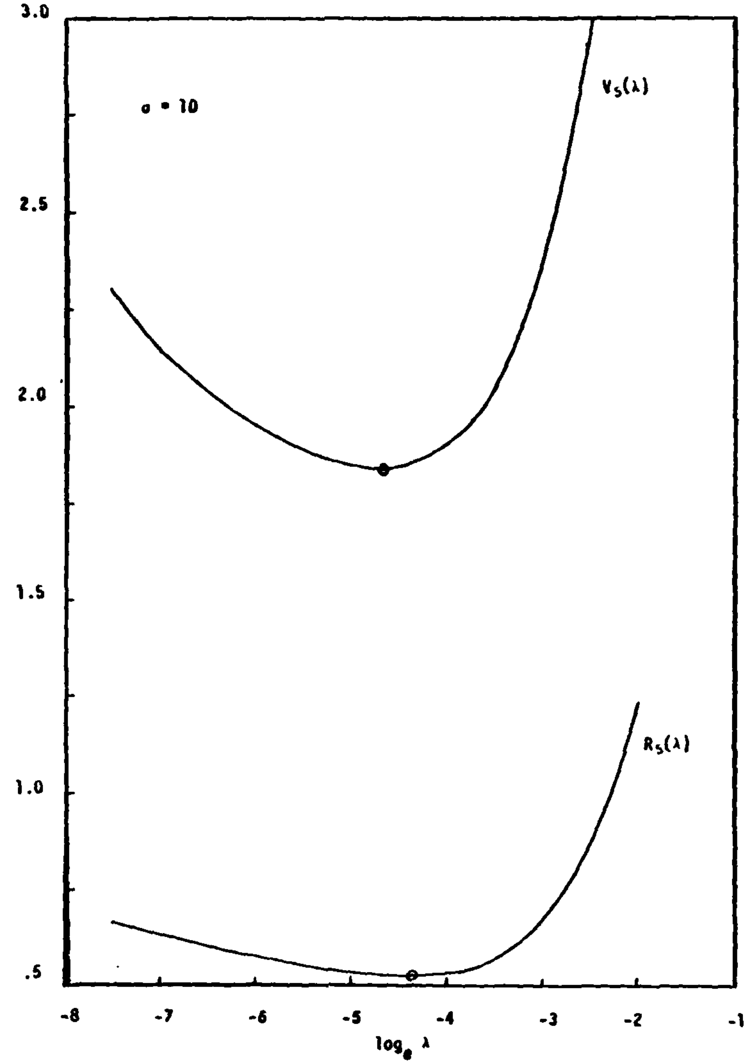
considered m = 2,3,4,5 and 6. The analyzed field is given by $u_{N,m,\lambda}$ of equation (2.8) where ξ_j is defined by (2.9), ϕ_ν is defined by (2.10), K is defined by (2.17) and (2.12), and T is defined by (2.16) and (2.19). For each m, $V_m(\lambda)$ is defined by (3.9), where D_σ is taken as the identity matrix since all measurement errors are assumed to have the same standard deviation. $V_m(\lambda)$ is computed as in Section 3, but since D_σ is the identity matrix then $\tilde{T} = T$. The earth was assumed "flat" and latitude and longitude coordinates were treated as (x,y) for the analysis of the field and then converted back to latitude and longitude in the contour maps given below.

In the first series of experiments we considered one field (to be called Example 1) and considered σ = 5,10,15, and 20 meters. For each data set (i.e. value of σ) we let m = 2,3,4,5 and 6. Let us first examine the choice of λ. In the first example discussed here, $\sigma = 10$ and $m = 5$. ($m = 5$ was the "estimated" m for this case, more about that next.) Figure 2 gives a plot of $V_5(\lambda)$ vs. λ and $R_5(\lambda)$. Here $R_m(\lambda)$ is defined as

$$R_m(\lambda) = \frac{1}{N}\sum_{i=1}^{N} (u_{n,m,\lambda}(t_i)-u(t_i))^2$$

where $u(t_i)$ is the "true", i.e. model 500mb height field at station i. Theoretically, $V_m(\lambda)$ should "track" $R_m(\lambda)$ near the minimum of $R_m(\lambda)$ (see Craven and Wahba (1979), Golub, Heath and Wahba (1977)). In practice $R_m(\lambda)$ is not known but in this example which is fairly typical, it can be seen that the minimizer, call

FIGURE 2



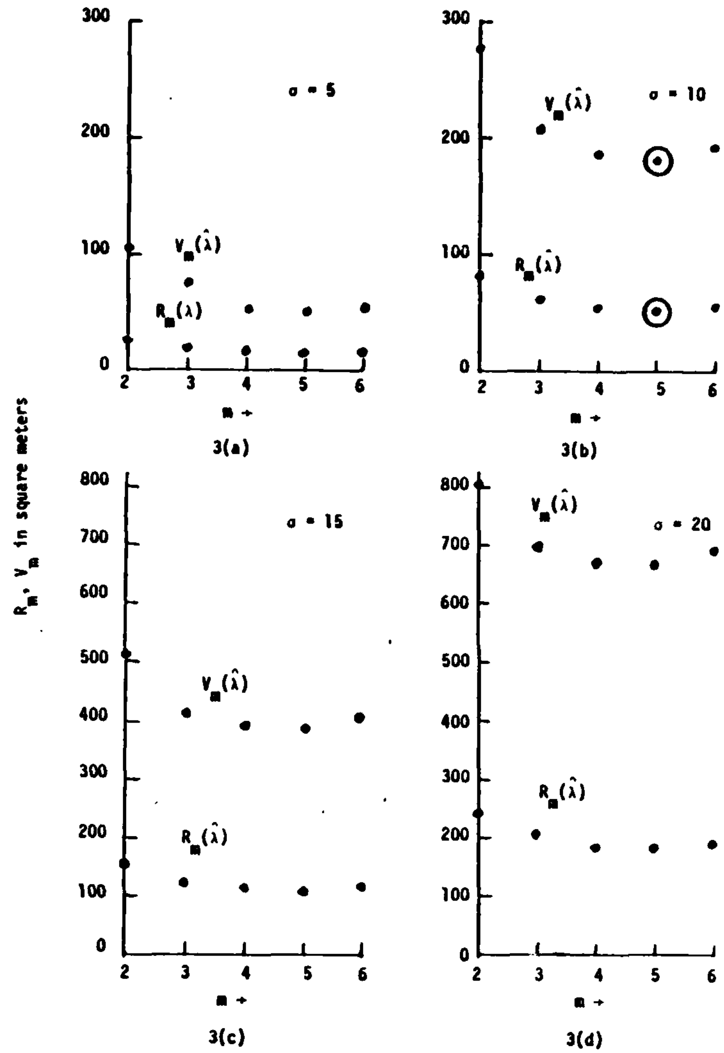
$R_m(\lambda)$ and $V_m(\lambda)$, m=5, Example 1

it $\hat{\lambda}$, of $V_m(\lambda)$ is a very good estimate of the minimizer of $R_m(\lambda)$. In fact the "inefficiency" $R_m(\hat{\lambda})/\min_\lambda R_m(\lambda) = 1.005$.

Figure 3 illustrates how m is chosen from the data, and how good this choice is. To study variability of the method with m and σ, the same set of 114 pseudo random numbers has been used in each of the twenty = 5 x 4 analyses behind Figure 3. The pseudo random number for station i was multiplied by σ = 5,10,15, and 20 in turn to get four data sets.

Figure 3(a) plots V_m at the minimizing value $\hat{\lambda}$ for m = 2,3,4,5 and 6 for the first data set (σ = 5). The minimizing value $\hat{\lambda}$ will be different in each case. According to Figure 3(a) the choice of m = 5 would be made from the data. For comparison $R_m(\hat{\lambda})$ is also given. Figures 3(b), 3(c) and 3(d) give the same plots for the other three data sets with σ = 10,15, and 20. It is seen that the choice m = 5 would be made from the data in each case. In general $R_m(\hat{\lambda})$ is very close to $\min_\lambda R_m(\lambda)$ and these plots suggest that choosing m to minimize $V_m(\hat{\lambda})$ will result in a good choice of m. However, $R_m(\hat{\lambda})$ for m = 4 and m = 6 is only slightly larger than $R_5(\hat{\lambda})$. The two points corresponding to the m = 5, σ = 10 case of Figure 2 are circled in Figure 3(b). Figures 4(a), (b), (c), and (d) give the model and analyzed field for m = 5 with the estimated λ for each σ tried. The model field contours (dashed lines) are the same in each figure. The analyzed field contours are solid lines. The contours are labeled in tens of meters.

FIGURE 3



$V_m(\hat{\lambda})$ and $R_m(\hat{\lambda})$

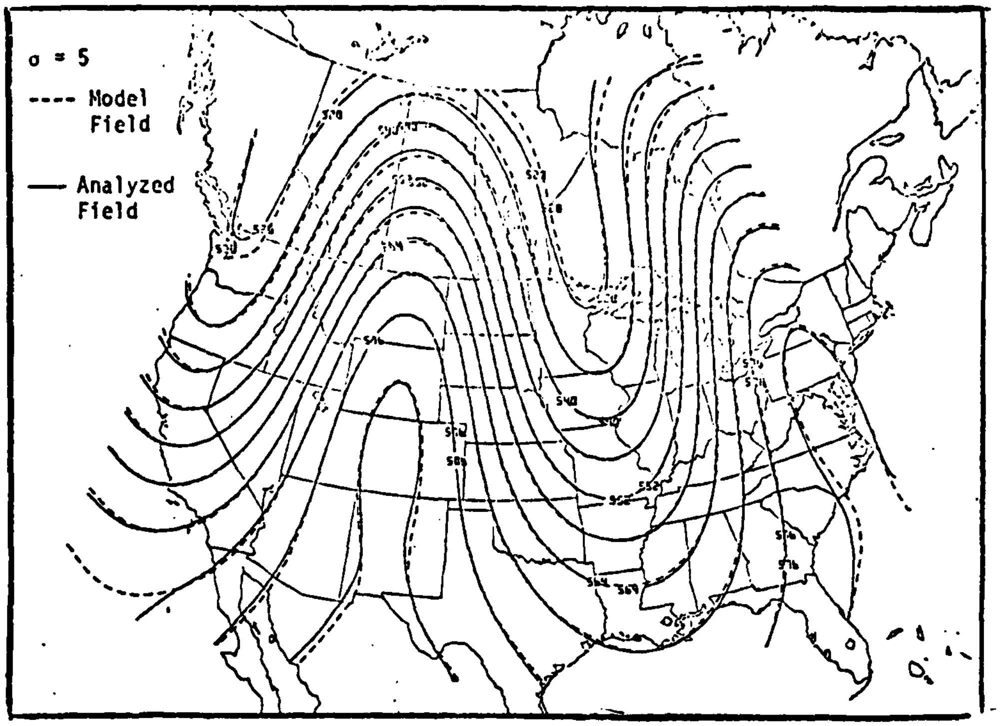


FIGURE 4a

Model and Analyzed Field, σ = 5

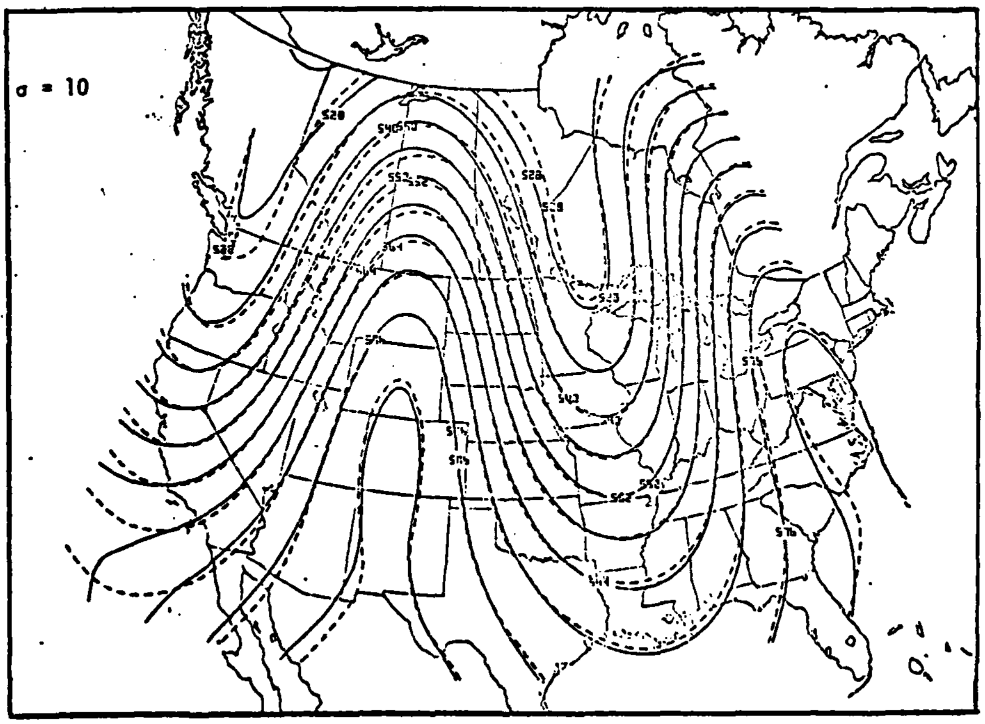


FIGURE 4b

Model and Analyzed Field, σ = 10

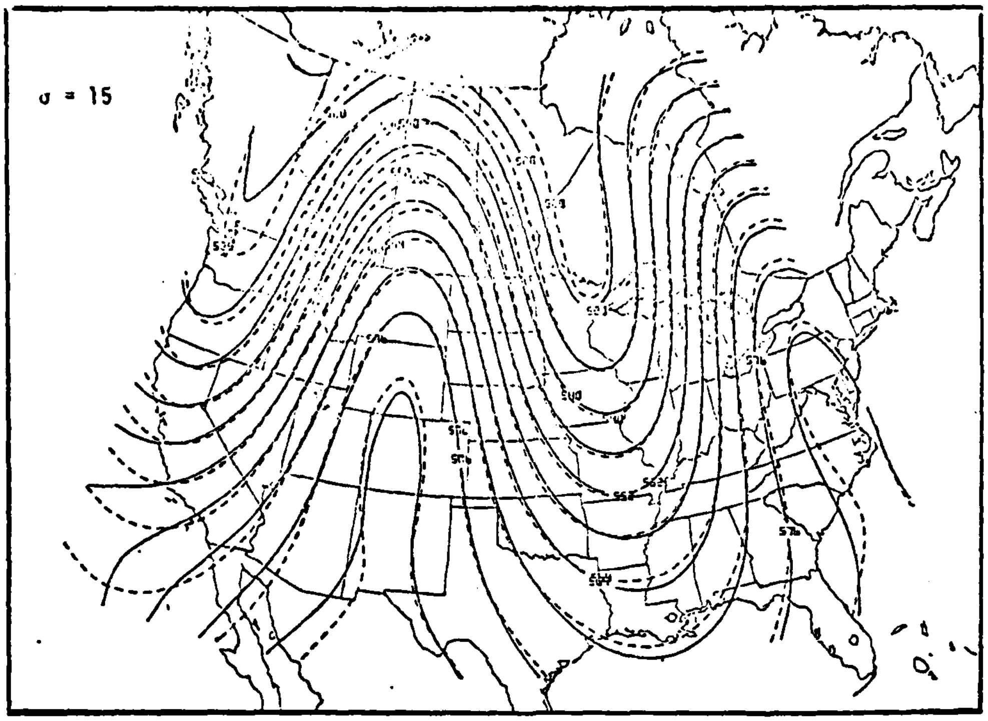


FIGURE 4c

Model and Analyzed Field, σ = 15

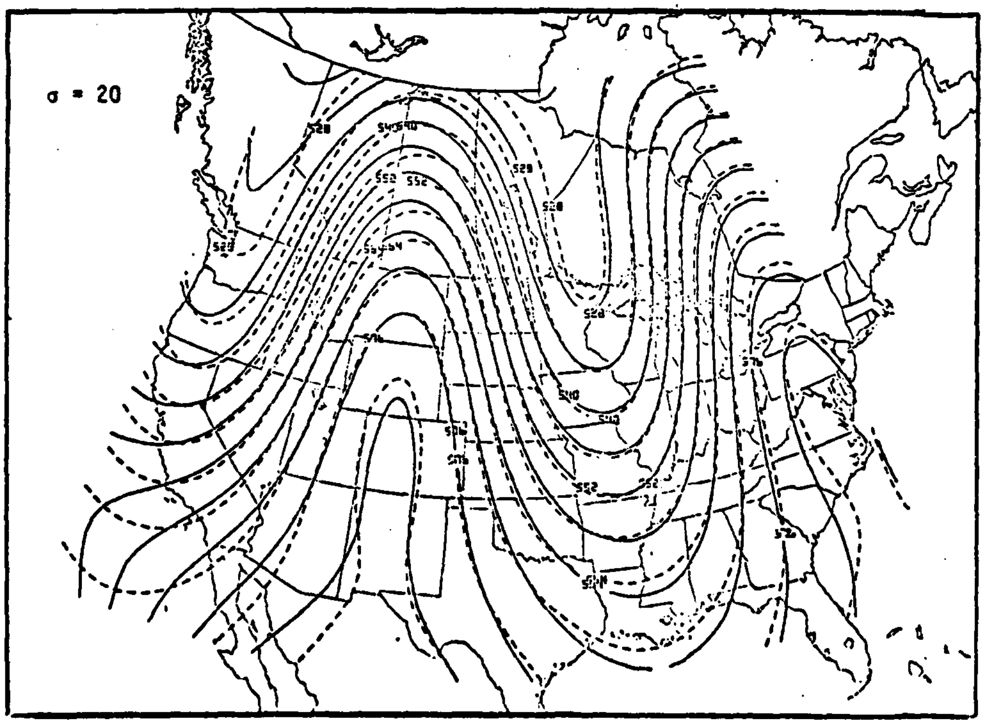


FIGURE 4d

Model and Analyzed Field, σ = 20

From the data behind Figure 3 one can establish that $(R_5(\hat{\lambda}))^{1/2}$ is between $.6\sigma$ and $.8\sigma$. Thus the measurement noise is being filtered out to give a better estimate over-all, of the station 500mb height than the measured heights!

In practice of course we want the analyzed field to be a good estimate of the true field over a whole region, not just at the points where it is measured. To determine how well this goal is being met the RMSE of the analyzed field over a 17×26 grid covering the region outlined over North America with a solid line in Figure 1 was computed. This RMSE is defined as follows:

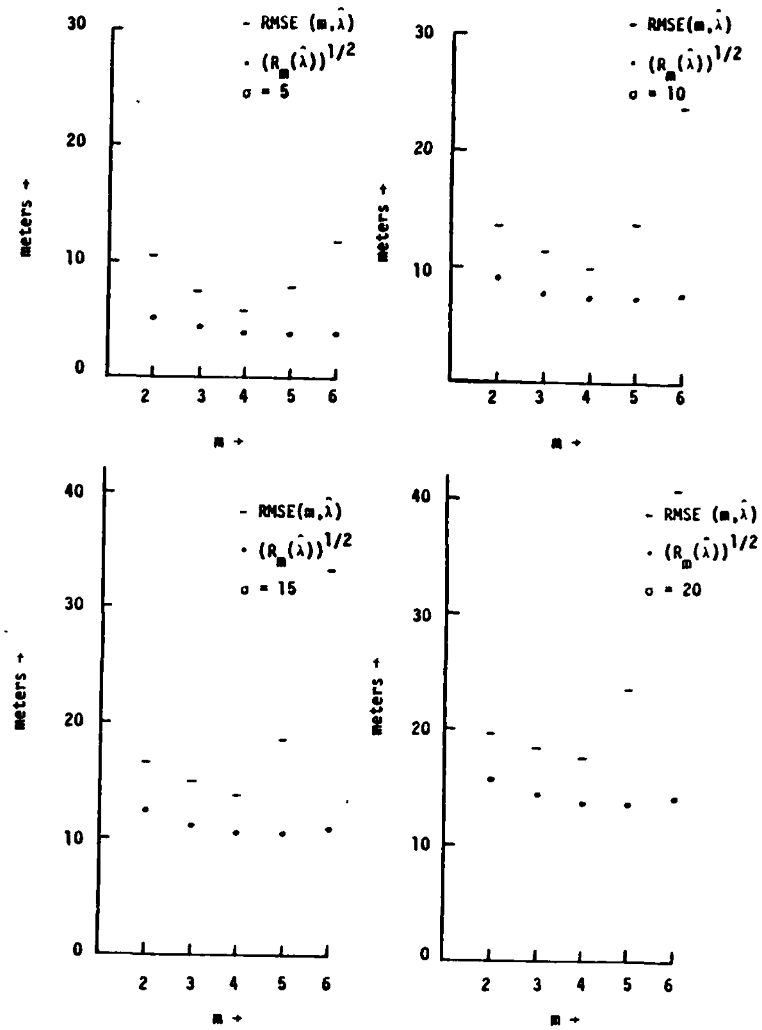
$$\text{RMSE} = \text{RMSE}(m,\hat{\lambda}) = \left(\frac{1}{17\times 26}\sum_{i=1}^{17}\sum_{j=1}^{26}[u_{114,m,\hat{\lambda}}(\theta_i,\phi_j)-u(\theta_i,\phi_j)]^2\right)^{1/2}$$

where $\hat{\lambda}$ is the estimated λ for each m. RMSE is, of course, an overall measure of how well an entire field can be estimated over a region from the 114 data points.

Figure 5 gives plots of $\text{RMSE}(m,\hat{\lambda})$ for the four values of σ tried. $\text{RMSE}(m,\hat{\lambda})$ is generally greater than $(R_m(\hat{\lambda}))^{1/2}$. For comparison $(R_m(\hat{\lambda}))^{1/2}$ is also plotted. The excess of $\text{RMSE}(m,\hat{\lambda})$ over $(R_m(\hat{\lambda}))^{1/2}$ reflects the inability of the method to interpolate between data points.

It can be seen from Figure 5 that by the RMSE criteria an m somewhat smaller than 5 would give slightly better results in these examples. To what extent this result on a model field carries over to real fields is really a question of how closely the model represents the real world with respect to the feature being tested.

FIGURE 5



$\text{RMSE}(m,\hat{\lambda})$ and $(R_m(\hat{\lambda}))^{1/2}$ vs. m for σ = 5,10,15 and 20

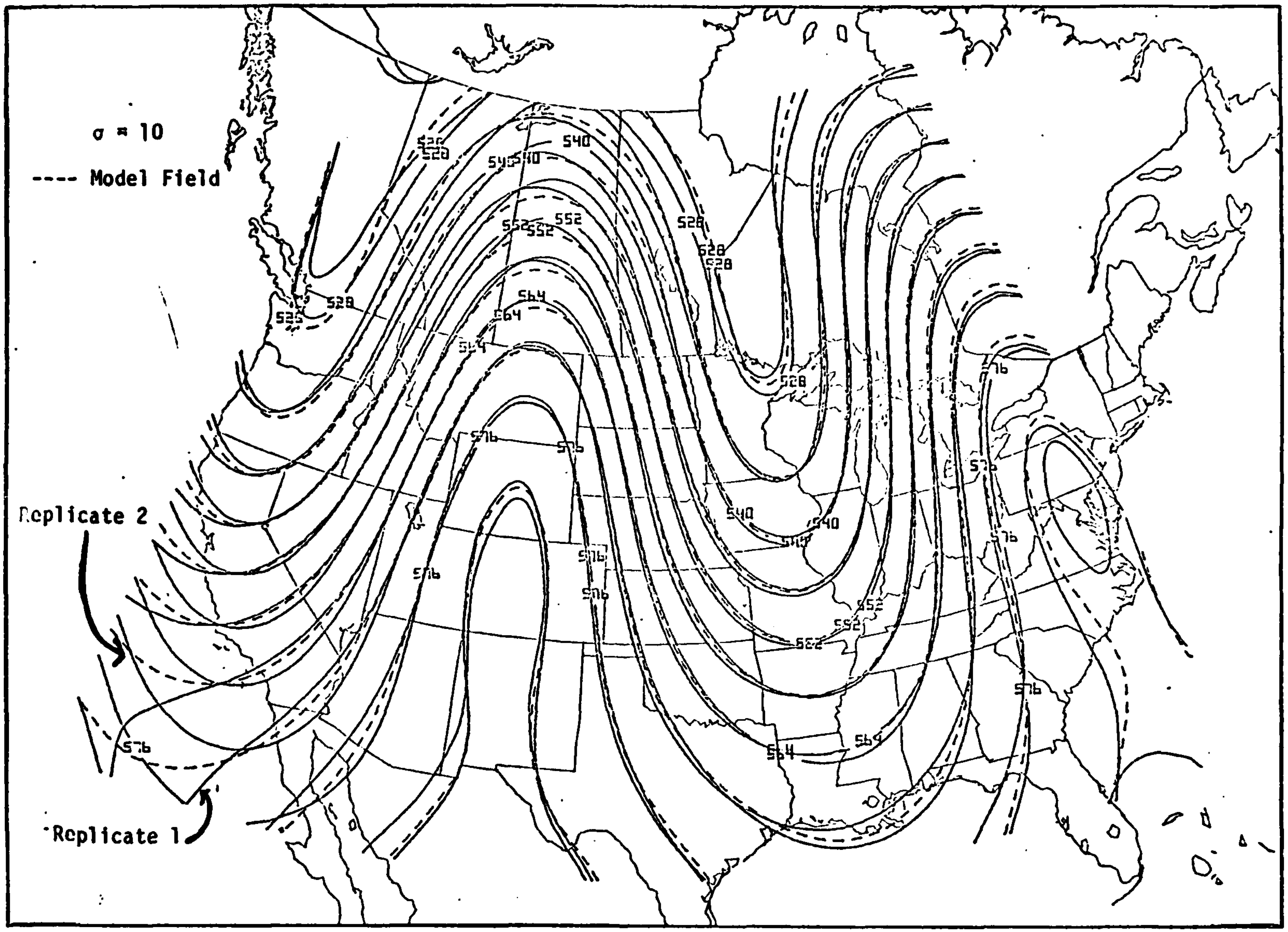
To get a feel for the variability of the analysis with actual variation in the measurment errors, Example 1 above with $\sigma = 10$ was replicated beginning with a new set of random numbers. $V_m(\lambda)$ was computed from the data and $m = 5$ was again chosen from the data.

The estimated value $\hat{\lambda}$ in the second replicate was very close to $\hat{\lambda}$ in the first replicate. (Remember that the "model" field is identical in both cases.) However, while the RMSE was 13.69 in the first replicate, it was 17.13 in this one. The model and the two analyzed fields for this case appear in Figure 6.

Finally, we look at variations as the field varied. Three other fields in addition to the first example were generated by moving the field from west to east. The four fields are characterized by the parameter $ALON_o$ in the model in Example 1, $ALON_o$=95, the other three cases are 90, 100 and 105. The second replicate with $ALON_o$ = 95 is used in this series, and the same set of 114 original random numbers used in the second replicate is used in the other three examples here. A set of data with $\sigma = 10$ was generated for each of these three new fields. The estimated values of m were

$ALON_o$	$\hat{m}$
90	6
95	5 (already given)
100	4
105	4

FIGURE 6



Model Field and Two Analyzed Fields with Replicated Data

Figure 7 gives a plot of the true and analyzed field in each of these four cases. The RMSE values were

ALON.	RMSE($\hat{m},\hat{\lambda}$)
105	8.40
100	10.80
95	17.13
90	13.08

We have used $\sigma = 10$ as a typical value here because the results in Thiebaux (1977) (Table I, first row, fifth column) suggest that the root mean square measurment error at Topeka is less than 10 meters (assuming 0 mean measurement errors). Note an earlier study (AWS 1955) has estimated σ at around 20 meters.

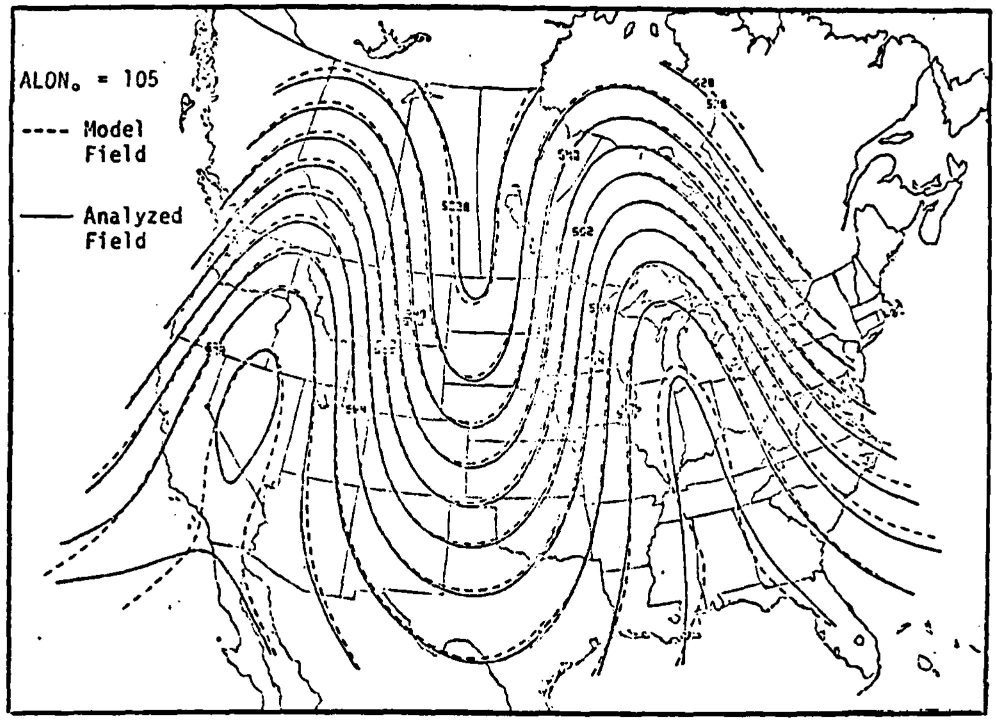
The question of whether in practice m and λ can be effectively chosen once and for all or should be estimated dynamically from the data has not been completely addressed here. This question can be addressed with "model" data only to the extent that the model represents the real world with respect to the phenomena being studied. Furthermore, if the criteria is minimum RMSE then this question cannot be answered with real data unless it is available on a fine grid. Predictive ability on the measurement grid can be studied in experiments philosophically like these of Thiebaux (1977), who omitted data from Tulsa and then examined how well the Tulsa data could be estimated from other data. We are presently doing this with both the isotropic and anisotropic method and preliminary results are very promising.

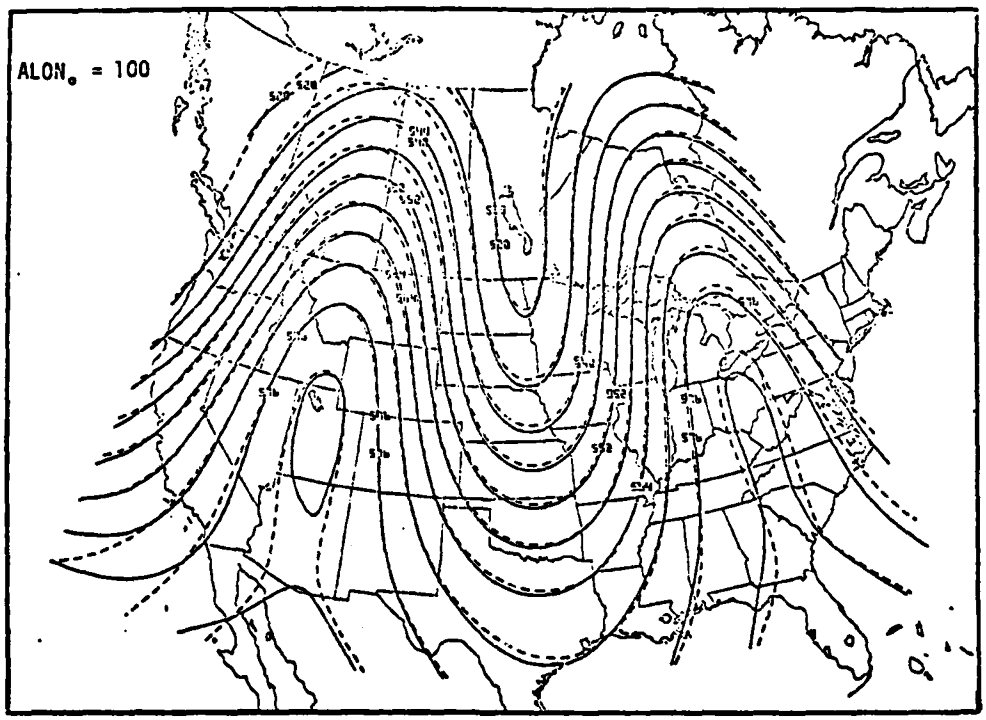
A few preliminary experiments we have carried out with a limited set of examples have resulted in effectively similar values of λ for fixed m. If m and λ can be fixed, then the cost of repetitive estimation of $u_{N,m,\lambda}$ from data from a given set of stations becomes very cheap.

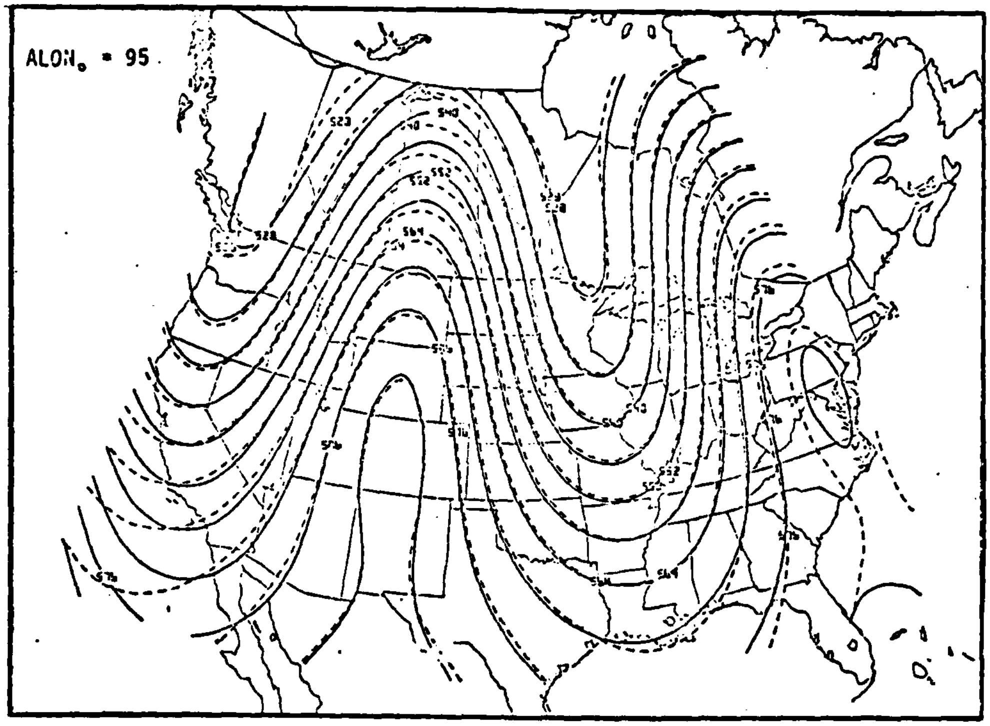
Ultimately questions whether m and λ should be estimated from the data or can safely be "fixed" at some prior value will have to be determined with respect to the ultimate use to which the analyzed field is put, e.g. if it is used in a forecast model, then one should determine whether dynamic estimation of λ and m is cost effective in terms of better forecasts.

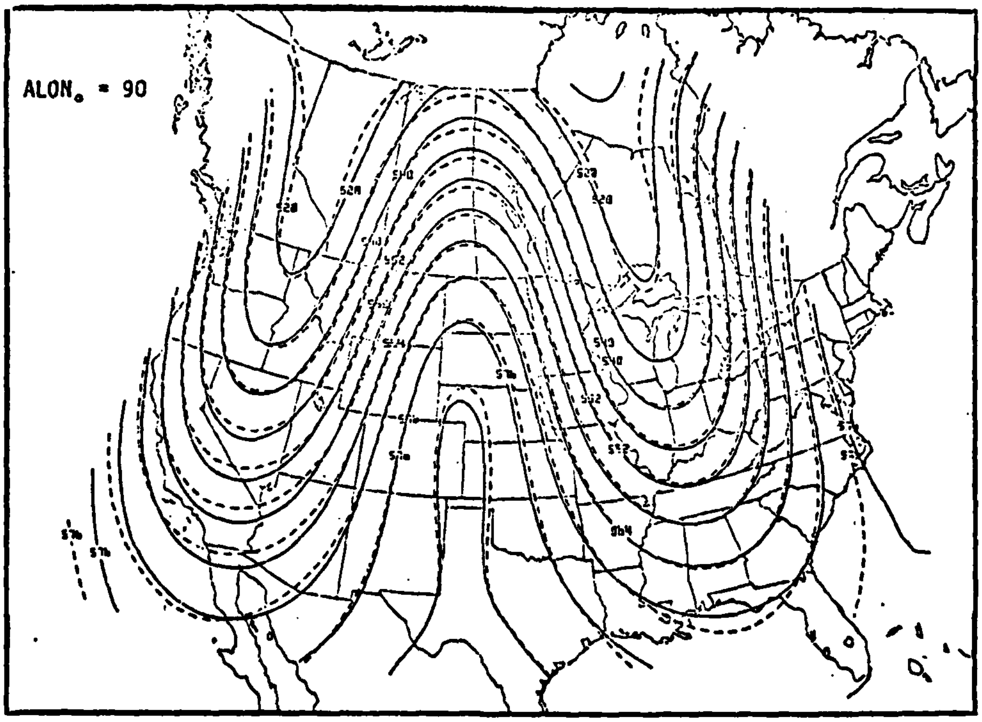
FIGURE 7

Four Examples with σ = 10






ALON₀ = 95


ALON₀ = 90

5. Acknowledgments

The authors wish to acknowledge a number of invaluable discussions with Professor Don Johnson and Drs. Tom Koehler and Tom Whittaker. This research was sponsored by the Atmospheric Sciences Section, National Science Foundation, under Grant No. ATM75-23223 and the Office of Naval Research under Contract No. N00014-77-C-0675.

Appendix 1

Mathematical Foundations

This appendix is not self-contained, to make it so would be a small monograph. Rather, it is an outline as to how known results may be combined to obtain the Main Theorem.

The appropriate rigorous definition of $\mathfrak{X}$ is: "$\mathfrak{X}$ is the vector space of all the (Schwartz) distributions for which all the partial derivatives in the distributional sense of total order m are square integrable." (Meinguet (1979) eqn. (4)) Choosing a set of points $\underset{\sim}{s}_1, \underset{\sim}{s}_2, \ldots, \underset{\sim}{s}_M$ in d space with the property

$$\sum_{j=1}^{M} \phi_\nu^2(\underset{\sim}{s}_j) > 0, \; \nu = 1,2,\ldots,M$$

$\mathfrak{X}$ can be decomposed into the direct sum of two spaces,

$$\mathfrak{X} = \mathfrak{X}_0 \oplus P_{m-1}$$

where P_{m-1} is the M dimensional space of polynomials of total degree less than m and $\mathfrak{X}_0$ is a proper Hilbert space of functions which are 0 at $\underset{\sim}{s}_1, \underset{\sim}{s}_2, \ldots, \underset{\sim}{s}_M$. It can be shown that $(J_m(u))^{1/2}$ is a norm on $\mathfrak{X}_0$ and

$$\|u\| = \left(\sum_{j=1}^{M} u^2(\underset{\sim}{s}_j)\right)^{1/2}$$

is a norm on P_{m-1}. Meinguet has found the reproducing kernel $K(\underset{\sim}{s},\underset{\sim}{t})$ for $\mathfrak{X}_0$ with the norm J_m and it is but a brief step to deduce from his results that the reproducing kernel $Q(\underset{\sim}{s},\underset{\sim}{t})$ for $\mathfrak{X}$ with

the norm given by (2.2) is

$$Q(\underset{\sim}{s},\underset{\sim}{t}) = K(\underset{\sim}{s},\underset{\sim}{t}) + P(\underset{\sim}{s},\underset{\sim}{t})$$

where

$$K(\underset{\sim}{s},\underset{\sim}{t}) = E_m(\underset{\sim}{s},\underset{\sim}{t}) - \sum_{\nu=1}^{M} p_\nu(\underset{\sim}{t})E_m(\underset{\sim}{s}_\nu,\underset{\sim}{s})$$

$$- \sum_{\mu=1}^{M} p_\mu(\underset{\sim}{s})E_m(\underset{\sim}{t},\underset{\sim}{s}_\mu)$$

$$+ \sum_{\mu,\nu=1}^{M} p_\mu(\underset{\sim}{s})p_\nu(\underset{\sim}{t})E_m(\underset{\sim}{s}_\mu,\underset{\sim}{s}_\nu)$$

$$P(\underset{\sim}{s},\underset{\sim}{t}) = \sum_{\nu=1}^{M} p_\nu(\underset{\sim}{s})p_\nu(\underset{\sim}{t})$$

and the p_ν are the M polynomials satisfying $p_\nu(\underset{\sim}{s}_j) = 1$, $\nu = j$, $= 0$ otherwise. See Aronszajn (1950) for a basic treatise on reproducing kernels. Using this reproducing kernel, it can be immediately deduced from Lemma (5.1) of Kimeldorf and Wahba (1971) that the solution to the minimization problem of (2.7) is given by

$$u_{N,m,\lambda}(\underset{\sim}{t}) = \sum_{j=1}^{N} c_j k_j(\underset{\sim}{t}) + \sum_{\nu=1}^{M} \bar{d}_\nu p_\nu(\underset{\sim}{t}) \qquad \text{(A.1)}$$

where

$$k_j(\underset{\sim}{t}) = L_{j(\underset{\sim}{s})}K(\underset{\sim}{s},\underset{\sim}{t})$$

and

$$\underset{\sim}{c} = M^{-1}(I-S(S'M^{-1}S)^{-1}S'M^{-1})\underset{\sim}{z} \qquad \text{(A.2)}$$

$$\bar{\underset{\sim}{d}} = (S'M^{-1}S)^{-1}S'M^{-1}\underset{\sim}{z} \qquad \text{(A.3)}$$

where

$$M = \tilde{K} + N\lambda D_\sigma^{\,2} \ ,$$

$\tilde{K}$ is the $N \times N$ matrix with jk^{th} entry

$$L_{j(\underset{\sim}{s})}L_{k(\underset{\sim}{t})}K(\underset{\sim}{s},\underset{\sim}{t})$$

and S is the $N \times N$ matrix with νj^{th} entry

$$L_j p_\nu \quad .$$

By premultiplying (A.2) by S' one obtains

$$S'\underset{\sim}{c} = 0$$

and by premultiplying (A.2) by M one obtains

$$N\underset{\sim}{c} = (\tilde{K}+N\lambda D_\sigma^2)\underset{\sim}{c} = \underset{\sim}{z} - S(S'M^{-1}S)^{-1}S'M^{-1}\underset{\sim}{z}$$

$$= \underset{\sim}{z} - S\tilde{\underset{\sim}{d}}.$$

It can then be checked that the basis $\{p_\nu\}_{\nu=1}^M$ for P_{m-1} can be replaced by the numerically more convenient basis $\{\phi_\nu\}_{\nu=1}^M$ by merely replacing S by the M x N matrix T with j^{th} entry $L_j\phi_\nu$ everywhere in the above. (This follows since $p_\mu = \sum_{\nu=1}^M \theta_{\mu\nu}\phi_\nu$ for some $\{\theta_{\mu\nu}\}$ and then $L_j p_\mu = \sum_{\nu=1}^M \theta_{\mu\nu} L_j \phi_\nu$). It can be checked that

$$\sum_{\nu=1}^M \tilde{d}_\nu p_\nu(\underset{\sim}{t}) = \sum_{\nu=1}^M d_\nu \phi_\nu(\underset{\sim}{t})$$

where

$$\underset{\sim}{d} = (T'M^{-1}T)^{-1}T'M^{-1}\underset{\sim}{z}$$

and that all equations involving $\underset{\sim}{c}$ still hold upon replacing S by T. The result is

$$u_{N,m,\lambda}(\underset{\sim}{t}) = \sum_{j=1}^N c_j k_j(\underset{\sim}{t}) + \sum_{\nu=1}^M d^*_\nu \phi_\nu(\underset{\sim}{t}) \qquad \text{(A.4)}$$

where

$$(\tilde{K} + N\lambda D_\sigma^2)\underset{\sim}{c} = \underset{\sim}{z} - T\underset{\sim}{d}^*$$

$$T'\underset{\sim}{c} = 0 \quad .$$

Now, let

$$\xi_j(\underset{\sim}{t}) = L_{j(\underset{\sim}{s})}E_m(\underset{\sim}{s},\underset{\sim}{t}) \quad .$$

Then it is easy to show that for any $\underset{\sim}{c}$ with $T'\underset{\sim}{c} = 0$, that

$$\sum_{j=1}^N c_j k_j(\underset{\sim}{t}) \equiv \sum_{j=1}^N c_j \xi_j(\underset{\sim}{t}) + (\text{a polynomial of degree} \le m-1). \qquad \text{(A.5)}$$

Using (A.5) and the fact that $T'\underset{\sim}{c} = 0$, one can show after some manipulation that $u_{N,m,\lambda}(\underset{\sim}{t})$ of (A.4) is equal to

$$u_{N,m,\lambda}(\underset{\sim}{t}) = \sum_{j=1}^N c_j\, \xi_j(\underset{\sim}{t}) + \sum_{\nu=1}^M d_\nu \phi_\nu(\underset{\sim}{t})$$

where

$$(K + N\lambda D_\sigma^2)\underset{\sim}{c} = \underset{\sim}{z} - T\underset{\sim}{d} \quad , \qquad \text{(A.6)}$$

$$T'\underset{\sim}{c} = 0 \quad , \qquad \text{(A.7)}$$

and K is the N x N matrix with jk^{th} entry $L_{j(\underset{\sim}{s})}L_{k(\underset{\sim}{t})}E_m(\underset{\sim}{s},\underset{\sim}{t})$. This is the Main Theorem.

We next obtain (3.10) and (3.11) from (2.13) and (2.14):

$$\underset{\sim}{c} = R(R'KR + N\lambda R'D_\sigma^{2}R)^{-1}R'\underset{\sim}{z} \qquad (3.10)$$

$$\underset{\sim}{d} = (T'D_\sigma^{-2}T)^{-1}T'D_\sigma^{-2}(\underset{\sim}{z} - K\underset{\sim}{c}) \qquad (3.11)$$

$$(K + N\lambda D_\sigma^{2})\underset{\sim}{c} + T\underset{\sim}{d} = \underset{\sim}{z} \qquad (2.13)$$

$$T'\underset{\sim}{c} = 0 \qquad (2.14)$$

Here R is any $N \times N\text{-}M$ matrix of rank $N\text{-}M$ satisfying $R'T = 0$. Since $T'\underset{\sim}{c} = 0$ there exists a unique $N\text{-}M$ vector γ, say, with

$$\underset{\sim}{c} = R\gamma \ . \qquad (A.8)$$

Left multiplying (2.13) by R' and substituting in (A.8) gives

$$R'(K + N\lambda D_\sigma^{2})R\gamma = R'\underset{\sim}{z}$$

$$\gamma = (R'(K + N\lambda D_\sigma^{2})R)^{-1}R'\underset{\sim}{z} \qquad (A.9)$$

and left multiplying (A.9) by R gives (3.10). To get (3.11) left multiplying (2.13) by $T'D_\sigma^{-2}$ to get

$$T'D_\sigma^{-2}K\underset{\sim}{c} + T'D_\sigma^{-2}T\underset{\sim}{d} = T'D_\sigma^{-2}\underset{\sim}{z}$$

left multiplying by $(T'D_\sigma^{-2}T)^{-1}$ gives (3.11).

To obtain (3.12)

$$I-A_m(\lambda) = N\lambda D_\sigma^{2}R(R'KR+N\lambda R'D_\sigma^{2}R)^{-1}R' \qquad (3.12)$$

it is necessary to know that

$$L_k\xi_j = L_{k(\underset{\sim}{s})}L_{j(\underset{\sim}{t})} \quad E_m(\underset{\sim}{t},\underset{\sim}{s}) \ .$$

This is not hard to check from the definitions. Then one has

$$L_k u_{N,m,\lambda} = \sum_{j=1}^{N} c_j L_k \xi_j + \sum_{\nu=1}^{M} d_\nu L_k \phi_\nu \quad , \ k = 1,2,\dots,N$$

or

$$\begin{pmatrix} L_1 u_{N,m,\lambda} \\ L_2 u_{N,m,\lambda} \\ \vdots \\ L_N u_{N,m,\lambda} \end{pmatrix} = K\underset{\sim}{c} + T\underset{\sim}{d}$$

and by the definition of $A_m(\lambda)$, we have

$$K\underset{\sim}{c} + T\underset{\sim}{d} = A_m(\lambda)\,\underset{\sim}{z} \ ,$$

Thus

$$(I-A_m(\lambda))\underset{\sim}{z} = \underset{\sim}{z} - K\underset{\sim}{c} - T\underset{\sim}{d} \ . \qquad (A.10)$$

But from (2.13),

$$\underset{\sim}{z} - K\underset{\sim}{c} - T\underset{\sim}{d} = N\lambda D_\sigma^{\ 2}\underset{\sim}{c} \ . \qquad (A.11)$$

Substituting (3.10) into (A.11) gives (3.12).

We now give a brief argument why the N-M x N-M matrix $B = R'KR$ is always strictly positive definite. Let K_* and R_* be the special cases of K and R when $L_k u = u(\underset{\sim}{t}_k)$. Duchon (1976b) has shown in this case that $R_*'K_*R_*$ is always strictly positive definite for any $N \geq M$. By using the fact that all continuous linear functionals in a reproducing kernel Hilbert space are limits of sums of evaluation functionals, one can show the positive definiteness in general. See Dyn and Wahba (1979) for more details.

We close this appendix with the observation that one can also enforce strong constraints by the same methods. Suppose one wishes to minimize

$$\frac{1}{N}\sum_{k=1}^{N_1} (L_k u - z_k)^2/\sigma_k^{\ 2} + \lambda J_m(u)$$

subject to

$$L_j u = z_j, \quad j = N_1 + 1, \ldots, N \ .$$

Explicit formulae (which are equivalent to letting $\sigma_j \to 0$, $j = N_1 + 1, \ldots, N$ in (3.10) and (3.11)) can be obtained by the same methods.

Appendix 2

500mb Height Model

As mentioned earlier, the height field used in the numerical experiments is the same as that used by T. Koehler. Koehler adopted the model of Sanders to represent meteorological phenomena of interest (in particular we used pressure surfaces) over an area the size of North America. In his model the height, z, of any pressure surface, p, at longitude, θ, and latitude, φ, is defined as follows:

$$z(\theta,\phi,p) = \hat{z}\cos((2\pi/L_\theta)(\theta_* - \theta + \Delta\theta))G'(\phi) + \bar{z} +$$
$$(T_m(1000)/\gamma)(1-(p/1000)^{(R\gamma/g)}) - (R/g)(\ln(1000/p) -$$
$$(\alpha/2)(\ln(1000/p))^2)((ar/\sin\phi_*)(\cos\phi_* - \cos\phi) + \hat{T}\cos((2\pi/L_\theta)(\theta_* - \theta))G(\phi))$$

where

$$\begin{aligned}
T_m(1000) &= 278K,\\
\hat{T} &= 10K,\\
\hat{z} &= 150m,\\
\bar{z} &= 90m,\\
R\gamma/g &= .0953,\\
a &= .9\times 10^{-5}\ {}^\circ K/m,\\
r &= 6371km,\\
\Delta\theta &= 9^\circ,\\
L_\theta &= 30^\circ,\\
\phi_* &= 45^\circ,\\
\alpha &= .621,\\
R &= 287.04m^2/s^2K,\\
g &= 9.8m/s^2,\\
p &= 500mb\\
\theta_* &= ALON_*;
\end{aligned}$$

$$G(\phi) = b[\frac{18}{\pi}(\phi-\phi_o)]^6 + c[\frac{18}{\pi}(\phi-\phi_o)]^4 + d[\frac{18}{\pi}(\phi-\phi_o)]^2 + e,$$

with

$$b = -1/60,$$
$$c = 11/60,$$
$$d = -40/60,$$
$$e = 1 ;$$

and

$$G'(\phi) = \int_{\phi_o}^{\phi} \frac{\sin \phi'}{\sin \phi_o} \frac{\partial G(\phi')}{\partial \phi'} d\phi' .$$

In the numerical experiments the parameter $ALON_o$ was varied taking the values 105,100,95 and 90. This parameter determined the longitude at which the wave "begins". Hence, by decreasing $ALON_o$ the wave "moves" from west to east. For the physical interpretation of the other constants and functional form of the model the reader is referred to Koehler (1979).

REFERENCES

Achtemeier, G.L. (1975). On the initialization problem: A variational adjustment method, Monthly Weather Review, 103,12,1089-1103.

Adams, R.A. (1975). Sobolev Spaces, Academic Press, New York.

Air Weather Service, (1955). Accuracies of Radiosonde Data, TR 105-133, Headquarters Air Weather Service Military Air Transport Service, USAF, Washington, D.C.

Akhiezer, N.I., and Glazman, I.M. (1961). Theory of Linear Operators in Hilbert Space, Vol. 1, Translated from the Russian by Merlind Nestell, Ungar, New York.

Aronszajn, N. (1950). Theory of Reproducing Kernels, Trans. Amer. Math. Soc., 68,337-404.

Craven, P., and Wahba, G., (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation, Numer. Math. 31, 377-403.

Duchon, Jean, (1976a). Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. R.A.I.R.O. Analyse numerique, 10, 12, pp. 5-12.

Duchon, Jean, (1976b). Fonctions spline du type "Plaque mince" en dimension 2. No. 231. Seminaire d'analyse numerique, Mathematiques Appliqueés, Universite Scientifique et Medicale de Grenoble.

Duchon, J., (1977). Erreur d'interpolation des fonctions de plusiers variables par les D^m splines. Seminaire d' Analyse Numerique, Mathematiques Appliqueés Universite Scientifique et Medicale de Grenoble, No. 268.

Dyn, N., and Wahba, G., (1979). On the estimation of functions of several variables from aggregated data, MRC Technical Summary Report #1974, Mathematics Research Center, University of Wisconsin-Madison, Madison, Wiscônsin.

Dyn, N., Wahba, G., and Wong, W.H., (1979). To appear J. Am. Stat. Soc. as comments on a paper by Tobler.

Fritz, S., et. al., (1972). Temperature sounding from satellites NOAA Technical Report NESS 59, U.S. Department of Commerce.

Golub, G., Heath, M., and Wahba, G., (1979). Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics, 21, 2, 215-223.

Kimeldorf, G., and Wahba, G., (1971). Some results on Tchebycheffian spline functions. J. Math. Anal. Applic., 33,82-95.

Koehler, T., (1979). A case study of height and temperature analysis derived from Nimbus-6 satellite soundings on a fine mesh model grid. University of Wisconsin-Madison, Department of Meteorology.

Kreiss, H.O., (1979a). Problems with different time scales for ordinary differential equations. University of Uppsala Computer Sciences Department, Technical Report #68.

Kreiss, H.O., (1979b). Problems with different time scales for partial differential equations. University of Uppsala Computer Sciences Department, Technical Report #75.

Lewis. J.M., (1972). An operational upper air analysis, Tellus XXIV, 6, 514-530.

Lewis, J.M., and Grayson, T.H., (1972). The adjustment of surface wind and pressure by Sasaki's variational matching technique. Journal of Applied Meteorology, 11, 586-597.

Meinguet, Jean, (1978). Multivariate interpolation at arbitrary points made simple. Report No. 118, Institute de Mathematique Pure et Appliqueé, Université Catholique de Louvain, to appear, ZAMP.

Meinguet, Jean, (1979). An intrinsic approach to multivariate spline interpolation at arbitrary points. To appear in the Proceedings of the NATO Advanced Study Institute on Polynomial and Spline Approximation, Calgary 1978, B. Sahney, Ed.

Paihua Montes , (1978). Quelques methodes numeriques pour le calcul de fonctions splines a une et plusiers variables. Thesis, Analyse Numerique Université Scientifique et Medicale de Grenoble.

Reinsch, C., (1967). Smoothing by spline functions. Numerische Mathematik 10, 177-183.

Sanders, F., (1971). Analytic solutions of the nonlinear omega and vorticity equation for a structually simple model of disturbance in the baroclinic westerlies. Monthly Weather Review, 99, 393-407.

Sasaki, Y., (1960). An objective analysis for determining initial conditions for the primitive equations. The A & M College of Texas, Department of Oceanography and Meteorology, TR 208.

Sasaki, Y., (1969). Proposed inclusion of time variation terms, observational and theoretical, in numerical variational objective analysis. Journal of the Meteorological Society of Japan, 47, 2, 115-124.

Sasaki, Y., (1971). A theoretical interpretation of anisotropically weighted smoothing on the basis of numerical variational analysis. Monthly Weather Review 99, 9, 698-708.

Smith, B.T., et. al., (1976). Matrix Eigensystem Routines-EISPACK Guide. Springer-Verlag.

Stephens, J.J., (1970). Variational initialization with the balance equation, Journal of Applied Meteorology, 9, 732-739.

Thiebaux, H.J., (1977). Extending estimation accuracy with anisotropic interpolation, Monthly Weather Review, 105, 6, 691-699.

Wahba, G. (1975). Smoothing noisy data with spline functions, Numer. Math., 24, 383-393.

Wahba, G. (1978). Automatic smoothing of the log spectral density. University of Wisconsin-Madison, Technical Report #536. To appear, J.A.S.A.

Wahba, G. (1979a). How to smooth curves and surfaces with splines and cross-validation, in "Proceedings of the 24th Design of Experiments Conference", F. Dressel, Ed., Academic Press.

Wahba, G. (1979b). Convergence rates of "Thin Plate" Splines when the data are noisy, University of Wisconsin-Madison, Department of Statistics, Technical Report #557, to appear in Proceedings of the Conference on Smoothing, T. Gasser and M. Rosenblatt, eds., Springer-Verlag, to appear.

Wahba, G. (1979c). Spline interpolation and smoothing on the sphere. University of Wisconsin-Madison, Department of Statistics Technical Report #534, to appear.

Wendelberger, J. (1980). Thesis, to appear.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report #578	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): SOME NEW MATHEMATICAL METHODS FOR VARIATIONAL OBJECTIVE ANALYSIS USING SPLINES AND CROSS-VALIDATION	5. TYPE OF REPORT & PERIOD COVERED: Scientific Interim
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Grace Wahba, James Wendelberger	8. CONTRACT OR GRANT NUMBER(s): N00014-77-C-0675
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, University of Wis., 1210 West Dayton Street, Madison, Wisconsin 53706	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, 800 North Quincy Street, Arlington, VA 22217	12. REPORT DATE: September 1979
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(See reverse side)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Let $\phi(x,y,p,t)$ be a meteorological field of interest, say height, temperature, a component of the wind field, etc. We suppose that data $\{\phi_i\}_{i=1}^N$ concerning the field of the form $\phi_i = L_i\phi + \epsilon_i$ are given, where each L_i is an arbitrary continuous linear functional and ϵ_i is a measurement error.

The data ϕ_i may be the result of theory, direct measurements, remote soundings, or a combination of these. We develop a new mathematical formalism exploiting the method of Generalized Cross Validation, and some recently developed optimization results, for analyzing this data. The analyzed field, $\phi_{n,m,\lambda}$, is the solution to the minimization problem: Find ϕ in a suitable space of functions to minimize

$$\frac{1}{N}\sum_{i=1}^{N}\frac{(L_i\phi-\phi_i)^2}{\sigma_i^2} + \lambda J_m(\phi) \qquad (*)$$

where

$$J_m(\phi) = \sum_{\alpha_1+\alpha_2+\alpha_3+\alpha_4=m} \frac{m!}{\alpha_1!\alpha_2!\alpha_3!\alpha_4!} \iiiint \left(\frac{\partial^m \phi}{\partial x^{\alpha_1}\partial y^{\alpha_2}\partial p^{\alpha_3}\partial t^{\alpha_4}}\right)^2 dx\,dy\,dp\,dt .$$

The σ_i^2 are assumed mean square errors. Functions of d = 1,2, or 3 of the four variables x,y,p,t are also considered. Under rather general conditions, we give an explicit representation for the minimizer of (*). The parameter λ controls the tradeoff between the infidelity of the analyzed field to the data, and the roughness of the analyzed field as measured by $J_m(\cdot)$.

Alternatively λ may be thought of as controlling the half-power point of the implied data filter. m controls the number of continuous derivatives that $\phi_{N,m,\lambda}$ will possess, alternatively, m may be thought of as controlling the steepness or "roll-off" of the data filter. High m corresponds to a steep roll-off. The parameters λ and m are chosen by the method of Generalized Cross Validation (GCV). This method estimates that λ and m for which the implied data filter has maximum predictive capability. This predictive capability is assessed by the GCV method by (implicitly) leaving out one data point at a time and determining how well the missing point can be predicted from the remaining data. The results extend those of Sasaki and others in several directions. In particular, no preliminary interpolation or smoothing of the data is required and it is not necessary to solve a boundary value problem or even assume boundary conditions to obtain a solution. Prior covariances are not assumed. The parameters λ and m play the role of signal to noise ratio and "order" of the covariance, these being the two most important parameters in the prior, and are estimated from the immediate data rather than historical data or guesswork. The numerical algorithm is surprisingly simple for any N with N^2 somewhat less that the high speed storage capacity of the computer.

The approach can be used to analyze temperature fields from radiosonde measured temperatures and satellite radiance measurements simultaneously, to incorporate the geostrophic wind approximation and other information. In a test of the method (for d = 2) simulated 500mb height data was obtained at discrete points corresponding to the U.S. radiosonde network, by using an

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

analytic representation of a 500mb wave and superimposing realistic random errors. The analytic representation was recovered on a fine grid with what appear to be impressive results.